ELSEVIER

Journal of Substance Abuse Treatment 35 (2008) 137–147

Journal of Substance Abuse Treatment

Regular article

Assessing fidelity in individual and family therapy for adolescent substance abuse

Aaron Hogue, (Ph.D.)[a,*], Sarah Dauber, (Ph.D.)[a], Priscilla Chinchilla, (M.A.)[a], Adam Fried, (M.A.)[a], Craig Henderson, (Ph.D.)[b], Jaime Inclan, (Ph.D.)[c], Robert H. Reiner, (Ph.D.)[d], Howard A. Liddle, (Ed.D.)[e,f]

[a]*The National Center on Addiction and Substance Abuse at Columbia University, New York, NY 10017-6706, USA*
[b]*Department of Psychology, Sam Houston State University, Huntsville, TX 7734, USA*
[c]*Roberto Clemente Family Guidance Center, New York University School of Medicine, New York, NY, USA*
[d]*Behavioral Associates, New York City, NY, USA*
[e]*Department of Epidemiology and Public Health, Center for Treatment Research on Adolescent Drug Abuse, University of Miami Miller School of Medicine, Miami, FL, USA*
[f]*Department of Psychology, Center for Treatment Research on Adolescent Drug Abuse, University of Miami Miller School of Medicine, Miami, FL, USA*

Received 14 May 2007; received in revised form 22 August 2007; accepted 2 September 2007

Abstract

This study introduces an observational measure of fidelity in evidence-based practices for adolescent substance abuse treatment. The Therapist Behavior Rating Scale—Competence (TBRS-C) measures adherence and competence in individual cognitive–behavioral therapy and multidimensional family therapy for adolescent substance abuse. The TBRS-C assesses fidelity to the core therapeutic goals of each approach and also contains global ratings of therapist competence. Study participants were 136 clinically referred adolescents and their families observed in 437 treatment sessions. The TBRS-C demonstrated strong interrater reliability for goal-specific ratings of treatment adherence, and modest reliability for goal-specific and global ratings of therapist competence, evidence of construct validity, and discriminant validity with an observational measure of therapeutic alliance. The utility of the TBRS-C for evaluating treatment fidelity in field settings is discussed. © 2008 Elsevier Inc. All rights reserved.

Keywords: Treatment fidelity; Therapist competence; Adolescent substance use; Cognitive–behavioral therapy; Family therapy

1. Introduction

Evaluating the feasibility and effectiveness of research-developed treatments for substance use disorders in usual care settings has become a national health care priority (Institute of Medicine, 2006). Substance abuse treatment programs are facing increased stakeholder demands for adherence to empirically based practice guidelines (Hayes, Barlow, & Nelson-Gray, 1999), and many programs describe their efforts to implement evidence-based practices as a cornerstone of quality care (e.g., Henderson et al., 2007; Mark et al., 2006). In addition, emerging research indicates that strong fidelity is critical for successful transportation of research-based protocols (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Henggeler, Pickrel, & Brondino, 1999). Reliable, brief, and cost-efficient evaluation tools are therefore needed to assess fidelity to multicomponent treatments that stand ready for dissemination to clinical settings (Carroll et al., 2000; Garland, Hurlburt, & Hawley, 2006).

There has been a recent surge in the number of empirically supported treatments for adolescent substance

* Corresponding author. Health and Treatment Research, National Center on Addiction and Substance Abuse at Columbia University, 19th Floor, 633 Third Avenue, New York, NY 10017, USA. Tel.: +1 212 841 5278; fax: +1 212 956 8020.
E-mail address: ahogue@casacolumbia.org (A. Hogue).

0740-5472/08/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsat.2007.09.002

abuse (for reviews, see Muck et al., 2001; Williams, Chang, & Addiction Centre Adolescent Research Group, 2000). However, only a handful of fidelity instruments have been developed to measure the implementation of evidence-based practices with adolescent drug users (e.g., Henggeler et al., 1999; Hogue et al., 1998), and these assess treatment adherence but not therapist competence. Treatment adherence generally refers to the quantity or extent of specific treatment techniques used in session, whereas therapist competence refers to the quality or skill with which interventions are delivered (Waltz, Addis, Koerner, & Jacobson, 1993). Elements of therapist competence include knowledge of client issues, appropriateness and timing of interventions, and degree of responsiveness to client in-session behaviors (Stiles, Honos-Webb, & Surko, 1998). This study introduces a measure of therapist competence for adolescent drug treatments that incorporates these important elements.

Competence ratings are typically based on observer reports, inasmuch as therapist reports of their own clinical proficiency do not match observer accounts (Levin, Owen, Stinchfield, Rabinowitz, & Pace, 1999; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004) and clients do not have sufficient expertise to judge treatment quality per se. In research studies with adult clinical populations, two methods have been commonly used to measure competence. In the global rating method, a single item ("How competent was the therapist in this session?") or a few interrelated items (e.g., therapist skill, empathy, and nonverbal behavior; Carroll, Connors, et al., 1998) are used to rate the observed portion of treatment. Advantages of this method include high face validity and relative ease in training judges; drawbacks include a lack of specificity in describing components of the particular model being assessed. In the discrete technique method, multiple intervention techniques considered to be signature therapeutic ingredients of a particular model are rated separately. In discrete technique fidelity scales, two separate ratings are given for each technique: a "quantity" score to capture adherence and a "quality" score to capture competence. Fidelity studies of adults with depression (Barber, Crits-Christoph, & Luborsky, 1996; Shaw et al., 1999) and substance use disorders (Barber, Foltz, Crits-Christoph, & Chittams, 2004; Carroll et al., 2000) have used this method. Advantages include the ability to discriminate adherence from competence and the high degree of specificity in assessing model components.

There are two limitations to using either the global rating method or the discrete technique method for measuring competence. First, these approaches do not closely approximate the theory-driven processes of case formulation and treatment planning that guide clinical practice, training, and supervision. Case formulation and treatment planning tend to revolve around molar therapeutic goals that embody the model's underlying principles of change and provide structure in clinical decision making for each client (Kazdin, 1999; Stiles et al., 1998). Therapeutic goals themselves comprise multiple integrated intervention techniques that extend across several sessions (Diamond & Diamond, 2002). Therapeutic goals are the clinical blueprint of a treatment model, whereas discrete techniques are the clinical tools. Second, global and discrete measures tend to focus on the behavior of the therapist. In clinical practice, however, competence is largely determined by the therapist's ability to adapt continually to developments in clients' lives as they occur in and out of session, while still adhering to the specified clinical framework (Stiles et al., 1998). Therapist responsiveness to client behaviors in time-referral problems, interpersonal manner, amenability to intervention efforts, and so forth is central to competent model delivery (Waltz et al., 1993).

To address these two limitations in competence assessment, we developed the Therapist Behavior Rating Scale—Competence (TBRS-C), an observational measure of adherence and competence in individual and family-based therapy for adolescent substance abuse. The TBRS-C focuses on molar therapeutic goals, rather than on specific techniques, and includes both goal-specific and global ratings of therapist competence. The TBRS-C also contains separate global ratings of therapist skill, responsiveness, and overall competence.

The TBRS was initially developed as a discrete technique adherence scale to assess fidelity in the same randomized trial of individual cognitive–behavioral therapy (CBT) versus multidimensional family therapy (MDFT) from which this study sample was drawn. Hogue et al. (1998) used the original TBRS to confirm basic treatment adherence and differentiation in the trial: CBT focused on antecedents and consequences of drug use and behavioral skills building, whereas MDFT focused on family interactions and systemic interventions. The current study expands upon Hogue et al. in two important ways. First, it examines fidelity to the core therapeutic goals of each model (five in CBT, four in MDFT) in lieu of 26 discrete techniques contained in the original TBRS. The molar therapeutic goals contained in the TBRS-C each comprise multiple discrete treatment techniques, including all those in the original TBRS and others. Second, this study includes a multidimensional assessment of therapist competence in addition to treatment adherence.

The primary aim of this study was to examine the interrater reliability, construct validity, and discriminant validity of the TBRS-C in measuring adherence and competence in CBT and MDFT for adolescent substance abuse. This study is among the first to assess competence and adherence in evidence-based treatments for adolescent drug problems. Because specific clinical expertise is needed to make valid assessments of competence in model delivery (Waltz et al., 1993), separate coding groups were recruited for each treatment condition (as in Barber et al., 2004): CBT-knowledgeable judges rated CBT sessions, and family-therapy-knowledgeable judges rated MDFT sessions. Interrater reliability and variance components (therapist, client,

and treatment-phase effects) were calculated for individual items and averaged scales. Correlations among TBRS-C adherence and competence scores were used to examine construct validity; correlations between the TBRS-C and an observational measure of therapeutic alliance were used to examine discriminant validity. A second aim of the study was to compare reliability and validity findings for goal-specific competence ratings versus global ratings of therapist skill, responsiveness, and overall competence.

2. Materials and methods

The study was conducted with approval by the governing Institutional Review Board. Active consent from caregivers and active assent from adolescents were collected in writing from all participants, and active consent to judge fidelity was collected from all study therapists.

2.1. Sample

2.1.1. Clients

The client sample comprised 136 urban substance-abusing adolescents drawn from a larger randomized trial (N = 224) comparing individual CBT and MDFT (Liddle, 2002a). The cases selected for this study (62 for CBT and 74 for MDFT) included all those that met the following criteria: completed a baseline assessment, had at least one videotaped therapy session, and completed at least one posttreatment assessment (for future fidelity-outcome studies). The sample was 81% male, with an average age of 15.5 years (SD = 1.3) and a range of 13–17 years. The ethnic composition was approximately 70% African American, 20% European American, and 10% Hispanic American. Half of the adolescents were living in one-parent households, 14% had both biologic parents, and 36% had various other compositions. Yearly household income was <US$10,000 for 29% of the sample. Most adolescents were enrolled in school (76%) and were on juvenile probation (63%) at intake, and 32% were court ordered to receive treatment. Eighty percent met *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised* (*DSM-III-R*) criteria for a substance use disorder (21% for alcohol abuse/dependence and 79% for marijuana abuse/dependence), 79% met *DSM-III-R* criteria for an externalizing disorder, and 49% met *DSM-III-R* criteria for an internalizing disorder.

2.1.2. Therapists

The nine therapists who delivered the treatments (four in CBT and five in MDFT) ranged in age from 29 to 54 years (M = 40). The CBT therapists (two female therapists) included two African Americans and two European Americans; one had a master's degree and three had doctoral degrees. MDFT therapists (three female therapists) included three African Americans and two European Americans; four had a master's degree and one had a doctoral degree.

2.1.3. Treatments

2.1.3.1. Individual CBT. The individual CBT model for multiproblem adolescent drug users (Turner, 1992; Waldron & Kaminer, 2004) is based on a broadly defined cognitive–behavioral framework that emphasizes a harm-reduction approach to substance use. Initial sessions focus on identifying and prioritizing adolescent problems and constructing the treatment contract. The intensive cognitive–behavioral treatment program focuses on increasing coping competence and reducing problematic behaviors using intervention modules tailored to the individual adolescent: health education, self-monitoring, problem solving and communication skills, identifying cognitive distortions, and increasing prosocial activities. Role rehearsal and homework assignments are utilized to reinforce new skills. Final sessions focus on relapse prevention and maintenance of gains.

2.1.3.2. Multidimensional family therapy. MDFT (Liddle, 2002b) is a family-based ecological treatment for adolescent drug abuse and related behavior problems. MDFT therapists work simultaneously in four interdependent treatment domains according to the particular risk and protection profile of the adolescent and family. The adolescent domain helps teens engage in treatment, communicate effectively with parents and other adults, and develop social competence and alternative behaviors to drug use. The parent domain engages parents in therapy, increases their behavioral and emotional involvement with the adolescents, and improves parental monitoring and limit setting. The interactional domain focuses on decreasing conflict and improving emotional attachments and patterns of communication and problem solving using multiparticipant family sessions. The extrafamilial domain fosters family competency and collaborative involvement within all social systems in which the teen participates (e.g., school, juvenile justice, and recreation). At various points throughout treatment, therapists meet alone with the adolescent, alone with the parent(s), or conjointly with the adolescent and parent(s), depending on the treatment domain and specific problem being addressed.

2.1.3.3. Treatment fidelity procedures. Therapists were given study cases after 4 months of training and upon achieving satisfactory levels of fidelity in pilot cases as judged by model developers. Therapists were supervised weekly by model experts via live individual supervision, videotape feedback, and group supervision. Both treatments prescribed office-based weekly sessions conducted for 16–24 weeks. For the study sample, cases completed an average of 12.3 sessions (SD = 8.7), with 59% of cases attending eight or more sessions.

2.1.4. Observational measures

2.1.4.1. Therapist Behavior Rating Scale—Competence. The TBRS-C is an observational measure of treatment adherence and therapist competence for CBT and MDFT

Table 1
Variance components and ICCs for TBRS-C ratings: CBT condition

Goals	Variance components				ICC
	Therapist	Client	Phase	Residual	
Adherence (N = 192)					
1. Establishing a Working Relationship	.00	.02	.82	.16	.83
2. Drug-Use Monitoring and Harm Reduction	.20	.25	.01	.54	.81
3. Behavioral Skills Training	.03	.04	.45	.48	.62
4. Cognitive Therapy Techniques	.01	.21	.24	.54	.73
5. Increasing Prosocial Behavior	.05	.23	.24	.49	.56
Goal average adherence	.14	.26	.00	.59	.74
Competence					
1. Establishing a Working Relationship (n = 90)	.12	.40	.00	.47	.48
2. Drug-Use Monitoring (n = 138)	.13	.50	.00	.37	.63
3. Behavioral Skills Training (n = 56)	.00	.50	.08	.42	.25
4. Cognitive Therapy Techniques (n = 30)	.00	.00	.05	.95	.01
5. Increasing Prosocial Behavior (n = 57)	.06	.39	.00	.55	.39
Goal average competence (n = 192)	.06	.45	.06	.42	.56
Global competence ratings (N = 192)					
Overall competence	.03	.36	.04	.57	.56
Skill	.02	.36	.03	.58	.49
Responsiveness	.00	.41	.05	.54	.49

whose format can be adapted for other manualized treatments of adolescent substance use. Scale items represent the core therapeutic goals of the given treatment model. Items are scored using a 7-point Likert-type scale with the following anchors: 1 = *not at all*; 3 = *somewhat*; 5 = *considerably*; 7 = *highly*. Each item receives a separate score for adherence and competence. Adherence ratings estimate the thoroughness with which goals are executed and the frequency with which they are addressed. Competence ratings estimate the technical quality of interventions (skillfulness) and their timing and appropriateness for the given client and situation (responsiveness). Each TBRS-C item assists judges in making competence assessments for the given therapeutic goal by describing important treatment context considerations (e.g., client interpersonal style and treatment phase), general keys to competent goal implementation, and guidelines for coding both adherence and competence. The measure is structured so that items are independent of one another but may co-occur in any given session segment. The TBRS-C also contains three global items for rating competence across the entire session: therapist skill, therapist responsiveness, and overall competence.

The TBRS-C describes five molar therapeutic goals for individual CBT (Table 1): Establishing a Working Relationship, Drug-Use Monitoring and Harm Reduction (exemplary techniques: analysis of drug-use behavior, refusal skills, and moderated use), Behavioral Skills Training (communication skills, decision making and problem solving, anger management, role playing, and relaxation training), Cognitive Therapy Techniques (cognitive monitoring and change strategies, and coping with drug-use thoughts), and Increasing Prosocial Behavior. It also contains four molar therapeutic goals for MDFT (Table 2): Adolescent Interventions (exemplary techniques: building and maintaining adolescent alliance, mapping ecological influences on prosocial and antisocial behavior, and exploring drug-use behaviors and consequences), Parent Interventions (building and maintaining parent alliance, reinforcing attachment and resuscitating hope, and enhancing parental monitoring and discipline), Family Interaction Interventions (meeting individually with family members to prepare for family sessions, resolving parent–adolescent impasses, and promoting positive family dialogue), and Extrafamilial Interventions (school and vocational interventions, and juvenile justice interventions).

2.1.4.2. Vanderbilt Therapeutic Alliance Scale—Revised. The Vanderbilt Therapeutic Alliance Scale—Revised (VTAS-R) is a revised 22-item version of the original VTAS (Hartley & Strupp, 1983) that defines therapeutic alliance as a collaborative and task-oriented relationship determined by client behaviors and therapist–client relationship. It has demonstrated strong interrater agreement (intraclass correlation coefficient [ICC] range, .80–.93) and internal consistency (Cronbach's α = .93–.96) in previous studies of alliance in family-based treatments for adolescent drug use (Diamond, Liddle, Hogue, & Dakof, 1999; Robbins, Turner, Alexander, & Perez, 2003; Shelef, Diamond, Diamond, & Liddle, 2005). In a previous study using the same clinical sample as this study, Hogue, Dauber, Faw, Cecero, and Liddle (2006) found that VTAS-R ratings yielded strong reliability and internal consistency in both conditions for mean scores averaged across the 22 items: $ICC_{(1,2)}$ = .90 and Cronbach's α = .98 for therapist–adolescent alliance in CBT; $ICC_{(1,2)}$ = .83 and α = .97 for therapist–adolescent alliance in MDFT; and $ICC_{(1,2)}$ = .62 and α = .98 for

Table 2
Variance components and ICCs for TBRS-C ratings: MDFT condition

Goals	Variance components				ICC
	Therapist	Client	Phase	Residual	
Adherence (N = 245)					
1. Adolescent Interventions	.00	.15	.09	.76	.73
2. Parent Interventions	.04	.15	.17	.64	.79
3. Family Interaction Interventions	.00	.22	.10	.68	.66
4. Extrafamilial Interventions	.00	.20	.12	.68	.64
Goal average adherence	.00	.28	.00	.72	.52
Competence					
1. Adolescent Interventions (n = 188)	.00	.36	.03	.60	.48
2. Parent Interventions (n = 189)	.13	.33	.00	.55	.48
3. Family Interaction Interventions (n = 109)	.00	.43	.00	.57	.29
4. Extrafamilial Interventions (n = 62)	.19	.22	.00	.59	.15
Goal average competence (n = 245)	.05	.43	.01	.51	.55
Global competence ratings (N = 245)					
Overall competence	.11	.43	.00	.46	.63
Skill	.14	.42	.00	.45	.53
Responsiveness	.11	.40	.00	.49	.56

therapist–parent alliance in MDFT. The mean scores for parent and adolescent alliance in MDFT were not significantly correlated (Pearson's $r = -.08$).

2.1.5. Sampling procedures

2.1.5.1. Therapist Behavior Rating Scale—Competence. Videotaped sessions were selected from Phase 1 of treatment (every study case) and from Phase 2 (when available). Phase 1 contained the first two available sessions between Sessions 1 and 5, so that judges could evaluate clients' presenting problems and early treatment developments as a context for coding later sessions. Phase 2 contained a randomly selected set of three consecutive sessions (when available) starting on Session 6. Identical sampling procedures were used for both conditions. However, fewer sessions from the CBT condition were included in this study due to its somewhat higher treatment dropout rate in the original clinical trial: 36% of cases randomized to CBT dropped from treatment prior to Session 6, compared to 31% in MDFT.

In CBT, 192 sessions were selected from 62 cases. Due to early treatment dropout, 36% of cases had Phase 1 tapes only. Across the 192 sessions, 54% were Phase 1 tapes, 29% were Phase 2 tapes that fell between Sessions 6 and 12, and 17% were Phase 2 tapes between Sessions 13 and 25. For Phase 1 sets, 62% contained the first two sessions of treatment, 20% contained the first session only because no other videotape was available, and 18% contained some other configuration. For Phase 2 sets, 54% contained three consecutive sessions, 21% contained two consecutive sessions, 21% contained one session only, and 4% contained some other configuration. In MDFT, 245 sessions were selected from 74 cases. Due to dropout, 34% of cases had Phase 1 tapes only. Across the 245 sessions, 51% were Phase 1 tapes, 29% were Phase 2 tapes between Sessions 6 and 12, and 20% were Phase 2 tapes between Sessions 13 and 25. For Phase 1 sets, 57% contained the first two sessions of treatment, 15% contained the first session only, and 28% contained some other configuration. For Phase 2 sets, 67% contained three consecutive sessions, 19% contained two consecutive sessions, 4% contained one session only, and 10% contained some other configuration. Fourteen percent of sessions were with the adolescent alone, 12% were with the parent(s) alone, and 74% were conjointly with the adolescent and the parent(s).

2.1.5.2. Vanderbilt Therapeutic Alliance Scale—Revised. The current study utilized VTAS-R mean ratings from Hogue, Dauber, Faw, et al. (2006). For CBT, only therapist–adolescent alliance was coded. For MDFT, judges completed separate alliance protocols while viewing the tape: one for the adolescent if present and one for the parent (or two, then averaged) if present. Due to resource limitations, only one session from Phase 1 (Session 2 for 69% of cases) and one session from Phase 2 (randomly selected) were coded. In CBT, 71 sessions (42 sessions in Phase 1; 29 sessions in Phase 2) across 47 cases were coded for adolescent alliance. In MDFT, rated were 73 sessions for adolescent alliance (47 sessions in Phase 1; 26 sessions in Phase 2) and 72 sessions for parent alliance (48 sessions in Phase 1; 24 sessions in Phase 2) across 67 cases. The total number of MDFT sessions coded was 93 (58 sessions in Phase 1; 35 sessions in Phase 2); most MDFT tapes (n = 52; 56%) received both adolescent and parent ratings because both members participated in a session for at least 20 minutes.

2.1.6. Observational coding procedures

2.1.6.1. Fidelity judges. Two coding groups were recruited. CBT judges (n = 7) were recruited from a private clinic specializing in cognitive–behavioral treatments for mental health disorders: three European American women, three

European American men, and one Caribbean man. CBT judges averaged 4.8 years (*SD* = 3.7) of postgraduate therapy experience and 4.0 years (*SD* = 3.4) of postgraduate CBT experience. MDFT judges (*n* = 8) were recruited from a community mental health clinic specializing in family-based treatment: two Hispanic American women, two Spanish women, one Hispanic–Asian American woman, one European American woman, and two Hispanic American men. They averaged 6.1 years (*SD* = 8.3) of postgraduate therapy experience and 4.9 years (*SD* = 7.9) of postgraduate experience in family therapy.

2.1.6.2. Fidelity ratings. Procedures used for training the CBT and MDFT coding groups were identical. Judges were trained during weekly 90-minute meetings for 4 months using a review of the respective coding manuals, in-group coding and a review of practice tapes, and exercises to increase understanding of scale items. Study coding commenced once both groups had reached a threshold reliability of ICC = .65 for the preponderance of items; thereafter, groups met biweekly for supportive training and monitoring of rater drift until coding had been completed. Judges rated all assigned sessions from a given case in chronological sequence for 2 weeks. Sessions were rated in their entirety, which ranged from 30 to 75 minutes and averaged 60 minutes. Two judges were assigned to code all tapes selected for each case, and judges were randomly paired with one another across the sample using a randomized block design (Fleiss, 1981).

Two judges rated each session, and pairs of ratings were averaged to create a final score for each scale item. Goal average scores were then created by averaging the final scores for the therapeutic goals (five items for CBT and four items for MDFT) for each session. Goal-specific competence ratings were not made for scale items that received a score of 1 (*not at all*) for adherence, which indicates nonoccurrence of that therapeutic goal in the observed session. Thus, judges rated the quality of a given item only when the therapist was observed to be working on that goal (Carroll et al., 2000). In contrast, the three global competence items were scored for every session.

Model-specific clinical expertise and instrument training are needed to code therapist competence in a valid manner (Waltz et al., 1993). It seemed counterintuitive to require judges clinically trained in CBT to make competence judgments about a model (MDFT) in which they were not experienced, and vice versa. Moreover, it would have been contrived to ask CBT judges to rate competence in achieving CBT treatment goals while viewing MDFT tapes, for which such goals would rarely be applicable. Thus, CBT sessions were rated using CBT items only, likewise for MDFT. In addition, TBRS-C items representing the main therapeutic goals of each model are theoretically independent of one another and are not intended to represent a correlated set of interventions (i.e., greater use of one goal is not correlated with greater use of another goal in any given session). For these reasons, internal consistency (Cronbach's alpha) was not calculated for either set of items.

2.1.7. Variance components analysis

Variance components analysis of the TBRS-C scores was conducted using mixed-models procedures in SAS. Variance components analysis involves partitioning the total variability among scores into reliable sources of variance (e.g., therapist, client, and so on). Along with regression coefficients, variance components are routinely estimated in random regression analytic techniques applied to nested research designs (e.g., mixed-model analysis of variance and hierarchical linear modeling). Variance components analysis was conducted on the goal average score for each item. Variance components were estimated using a restricted maximum-likelihood estimation method for the following sources of variance: Therapist, Client (nested within therapist), Phase, and Error. Each term was entered as a random effect in the analysis, and the estimates of variance were transformed into proportions of variance based on the estimates of the total variance. Because two individual judges were randomly assigned to each tape (rather than consistent pairs of judges being assigned, as in Barber et al., 2004), it was not possible to estimate a judge effect (i.e., the variance component of the averaged scores that was associated with judges).

3. Results

3.1. Reliability: Interrater reliability of goal-specific and global ratings

Both versions of the TBRS-C demonstrated good to excellent interrater reliability for adherence, but only fair to poor reliability for competence as measured by the ICC ($ICC_{(1,2)}$; Shrout & Fleiss, 1979). Cicchetti's (1994) criteria for classifying the utility of ICC magnitudes are as follows: <.40 = *poor*; .40–.59 = *fair*; .60–.74 = *good*; .75–1.00 = *excellent*. Table 1 contains ICC data for the CBT condition. The reliability of adherence ratings for the five specific goals of CBT ranged from .56 to .83. The reliability of competence ratings was generally much lower, ranging from .01 to .63. For the goal average scores, ICC = .74 for adherence and ICC = .56 for competence. Table 1 also contains the proportion of variance in ICC scores attributable to Therapist, Client, Phase, and Residual (Error). In four of the five scale items, the largest proportion of nonerror variance in adherence was attributed to Phase, indicating that these interventions were utilized, to a greater extent, in either one treatment phase or the other. In contrast, competence scores varied little between early and later treatment sessions. In addition, Client accounted for more variance than Therapist in all fidelity scores; the difference was especially pronounced for competence.

Table 3
Bivariate correlations among TBRS-C adherence and competence ratings: CBT condition

Goals	1	2	3	4	5	6	7	8	9
Adherence									
1. Establishing a Working Relationship	–	–	–	–	–	–	–	–	–
2. Drug-Use Monitoring and Harm Reduction	−.09	–	–	–	–	–	–	–	–
3. Behavioral Skills Training	−.50 **	.15 *	–	–	–	–	–	–	–
4. Cognitive Therapy Techniques	−.32 **	.25 **	.44 **	–	–	–	–	–	–
5. Increasing Prosocial Behavior	−.35 **	.11	.37 **	.02	–	–	–	–	–
Competence									
6. Establishing a Working Relationship	.61 **	.21	.09	.01	.16	–	–	–	–
7. Drug-Use Monitoring and Harm Reduction	.19 *	.74 **	.01	.20 *	−.12	.50 **	–	–	–
8. Behavioral Skills Training	−.01	.24	.68 **	.51 **	.11	−.03	.41 **	–	–
9. Cognitive Therapy Techniques	.08	.19	.32	.67 **	−.17	.92 *	.28	.48 *	–
10. Increasing Prosocial Behavior	−.01	.22	.26 *	.28 *	.65 **	.58 *	.56 **	.62 **	.79 **

* $p < .05$.
** $p < .01$.

Reliability data for MDFT are contained in Table 2. Adherence ratings were good to excellent, ranging from ICC = .64 to ICC = .79 for the four main goals. However, the reliability of competence ratings was again only fair to poor, ranging from .15 to .48. For the goal average scores, ICC = .52 for adherence and ICC = .55 for competence. As with CBT, Client accounted for greater proportions of variance in fidelity scores than Therapist.

Interrater reliability for global ratings of therapist competence was in the range of fair to good in each condition. For CBT (Table 1), ICC = .56 for overall competence, ICC = .49 for skill, and ICC = .49 for responsiveness. For MDFT (Table 2), ICC = .63 for overall competence, ICC = .53 for skill, and ICC = .56 for responsiveness. For both conditions, the reliability of the global ratings for each session compared favorably to the reliability of averaged ratings of individual therapeutic goals for each session.

3.2. Construct validity: Interitem correlations among TBRS-C items

Interitem correlations among the CBT items (Table 3) were mainly in the expected direction and support the construct validity of the scale. Adherence ratings for Establishing a Working Relationship were negatively correlated with the other four goals, as this goal is likely to be practiced during early treatment sessions in lieu of skills-oriented goals emphasized later on. Adherence scores for two skills-oriented items, Behavioral Skills Training and Cognitive Therapy Techniques, were positively correlated, $r(192) = .44$, $p < .01$. In addition, competence ratings for all five goals were positively correlated with one another (with one exception), indicating that therapists judged to be working competently on any one goal were seen as being generally competent across the board. In MDFT, adherence ratings for the four goals were weakly or negatively correlated with one another (Table 4), in keeping with their status as largely independent domains of therapeutic focus in this model. Competence ratings among the four MDFT goals were all positively correlated.

The pattern of correlations among goal average adherence and competence scores and the global rating of overall competence (Table 5) also supports the construct validity of the TBRS-C for both conditions. For CBT, goal average adherence and competence were moderately associated, $r(184) = .42$, $p < .01$, suggesting that adherence and competence were scored as related but distinct constructs. As expected, goal average competence was highly correlated

Table 4
Bivariate correlations among TBRS-C adherence and competence ratings: MDFT condition

Goals	1	2	3	4	5	6	7
Adherence							
1. Adolescent Interventions	–	–	–	–	–	–	–
2. Parent Interventions	−.31 **	–	–	–	–	–	–
3. Family Interaction Interventions	−.22 **	.01	–	–	–	–	–
4. Extrafamilial Interventions	.15 *	−.22 **	.07	–	–	–	–
Competence							
5. Adolescent Interventions	.42 **	.03	−.01	−.03	–	–	–
6. Parent Interventions	.01	.43 **	.10	−.19 **	.59 **	–	–
7. Family Interaction Interventions	−.12	.14	.57 **	−.16	.55 **	.64 **	–
8. Extrafamilial Interventions	.20	−.15	−.10	.48 **	.29 *	.16	.05

* $p < .05$.
** $p < .01$.

Table 5
Intercorrelations among TBRS-C adherence, TBRS-C competence, and VTAS-R alliance ratings for CBT and MDFT

	CBT			
Parameter	Goal average competence	Overall competence	Alliance: adolescent	
1. Goal average: adherence	.42 ** (184)	.50 ** (192)	.28 * (71)	
2. Goal average: competence		.68 ** (184)	.13 (67)	
3. Global rating: overall competence			.31 ** (71)	
	MDFT			
Parameter	Goal average competence	Overall competence	Alliance: adolescent	Alliance: parent
1. Goal average: adherence	.17 ** (245)	.23 ** (245)	.19 (73)	.16 (72)
2. Goal average: competence		.79 *** (245)	.40 ** (73)	.09 (72)
3. Global rating: overall competence			.36 ** (73)	.10 (72)
4. Therapeutic alliance: adolescent				−.04 (52)

* $p < .05$.
** $p < .01$.
*** $p < .001$.

with overall competence, $r(184) = .68$, $p < .01$. The two dimensions of overall competence, skill and responsiveness, were highly correlated also, $r(192) = .82$, $p < .001$ (not depicted in the table). Similarly in MDFT, there was a weak correlation between goal average adherence and competence, $r(245) = .17$, $p < .01$; a strong correlation between goal average competence and overall competence, $r(245] = .79$, $p < .001$; and a strong correlation between skill and responsiveness, $r(245) = .85$, $p < .001$ (not depicted).

3.3. Discriminant validity: Correlations between the TBRS-C and a measure of therapeutic alliance

Discriminant validity was examined by comparing TBRS-C fidelity ratings to independent ratings of therapeutic alliance from the VTAS-R (Table 5). In CBT, adolescent alliance was correlated with goal average adherence, $r(71) = .28$, $p < .05$, and overall competence, $r(71) = .31$, $p < .01$, but was not related to goal average competence. In MDFT, adolescent alliance was associated with both goal average competence, $r(73) = .40$, $p < .01$, and overall competence, $r(73) = .36$, $p < .01$. Parent alliance in MDFT was not significantly related to any fidelity score. The small (nonsignificant) to medium effect sizes observed in these fidelity–alliance correlations indicate that TBRS-C ratings were suitably distinct from concurrent judgments about the therapeutic relationship. In addition, findings generally converge with clinical expectations that greater competence begets stronger alliances, although this held true for adolescents only and not for parents. Other studies that examine correlations between fidelity and alliance have reported similar findings (e.g., Barber & Crits-Christoph, 1996; Carroll et al., 2000).

4. Discussion

This study presents initial reliability and validity findings for the TBRS-C, an observational measure of treatment fidelity applied to two empirically supported treatments for adolescent drug abuse, and individual CBT and MDFT. For both treatment models, the TBRS-C exhibited strong interrater reliability for adherence items, fair to poor reliability for individual competence items, but sufficient reliability for global competence ratings, and patterns of correlations among items that support construct validity. Ratings of adherence and competence were distinct from one another and from independent observations of a related treatment process, therapeutic alliance. Variance in adherence, especially competence scores, was attributable more to client effects than to therapist effects. These results demonstrate that fidelity measures focusing on molar therapeutic goals and incorporating therapist responsiveness into the assessment of competence can demonstrate acceptable psychometric properties to complement a high level of clinical validity.

Interrater reliability for adherence items of the TBRS-C was equivalent or superior to that reported for individual items on discrete scales (e.g., Barber, Liese, & Abrams, 2003; Morgenstern, Morgan, McCrady, Keller, & Carroll, 2001) and also for averaged scores comprising multiple items (e.g., Carroll, Connors, et al., 1998). The strong showing for adherence ICCs is noteworthy given that TBRS-C reliabilities were generated using a within-condition analytic strategy, whereby judges used CBT items to rate CBT tapes only (likewise for MDFT). This is a conservative approach to estimating reliability for adherence measures that attenuates the magnitude of intraclass correlations (Startup & Shapiro, 1993).

The interrater reliability of the competence ratings for individual therapeutic goals was generally weak and well below the magnitude found for competence items on most discrete techniques scales (e.g., Barber et al., 2003). The reliabilities of the global competence ratings (.56 for CBT and .63 for MDFT) and the average competence rating across therapeutic goals (.56 for CBT and .55 for MDFT) were modest but in keeping with the magnitude of competence

ratings in some studies (e.g., Barber & Crits-Christoph, 1996; James, Blackburn, Milne, & Reichfelt, 2001), although decidedly lower than in others (e.g., Carroll et al., 2000; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005). All would agree that it is easier to achieve reliable judgments about the quantity of a well-specified intervention than about its quality. The reliability gap between adherence and competence ratings is expectably even wider for a fidelity measure that demands judgments about molar multicomponent intervention goals. Finally, competence ratings were especially vulnerable to rater drift, wherein prestudy levels of acceptable reliability for individual items declined in the course of the study despite the occurrence of coder meetings throughout. This did not occur for adherence. It appears that continual exposure to new clients and repeated exposure to a small group of therapists engendered a more cohesive mindset among judges regarding how extensively a therapeutic goal was implemented, but more diverse opinions about how well.

The moderate correlations between averaged adherence and competence scores for both CBT (r = .42) and MDFT (r = .17) suggest that these constructs are related but not redundant, which can be considered a strength of the scale. These results compare favorably to adherence–competence correlations reported for discrete technique measures of drug counseling for cocaine abuse (.58; Barber et al., 1996; Barber et al., 2006), CBT for cocaine abuse (.96; Barber et al., 2003), and three treatments for co-occurring cocaine and alcohol problems (range, .21–.62; Carroll et al., 2000). For fidelity measures such as the TBRS-C to be maximally useful in field settings, they need to provide reliable and distinguishable data on both the kind of interventions delivered and the quality of treatment (Garland et al., 2006).

Greater variability in treatment adherence and competence was associated with clients than with therapists. The absence of therapist effects indicates that therapists were not consistently different from one another in fidelity across clients. One caveat to this finding is that all study therapists were intensively trained and deemed competent in model implementation prior to treating study cases, which reduced the potential spread of fidelity scores among therapists. This is highly desirable in controlled efficacy research but not likely to be found in real-world clinical settings. In contrast, the relatively strong client effects indicate that the TBRS-C was sensitive to client-level variations in fidelity that are reasonably expected to emerge across a large group of adolescents and families with differences in problem severity, clinical complexity, therapist–client fit, and the like. Strong client effects also suggest that fidelity was not a therapist-centered characteristic (i.e., therapists were not consistently adherent or competent across clients on their caseloads). Thus, in this small sample of research-trained providers, fidelity was a trait of therapist–client pairing, not of the therapist.

This study has an important methodological limitation with regard to competence evaluation. Judges viewed only a small number of sessions (between one and three sessions) selected from the later phase of treatment. Judges who do not observe (most) every session are not able to track the clinical progress of the case across treatment, which hampers their ability to provide fully informed case-specific assessments of competence (Elkin, 1999). For this reason, Waltz et al. (1993) argue that competence evaluations are better served by including fewer cases for which all sessions can be rated in sequence. Certainly, the TBRS-C can be used in this manner, resources permitting. In addition, the size of bivariate correlations involving competence scores may have been limited by their modest reliability; as a counterbalance, correlations among adherence and competence scores may have been inflated by common source and method variance.

4.1. Utility of the TBRS-C for fidelity evaluation in field settings

The TBRS-C appears suitable for meeting fidelity assessment needs in standard clinical practice. In controlled efficacy and effectiveness research, intensive and sustained involvement by model experts in training and supervising line staff is the norm (Schoenwald & Hoagwood, 2001). In the absence of this, quality assurance tools that document fidelity in real-world clinical conditions with precision and clinical sophistication are urgently needed. Whereas this study employed resource-intensive methods in a research setting to establish initial psychometric properties, in theory, the TBRS-C could be used as a supervision supplement, quality assurance measure, or both in field settings. However, further development work is required to verify that observational fidelity instruments are reliable and valid when used as self-report tools by frontline clinical supervisors and perhaps clinicians themselves (Carroll, Nich, & Rounsaaville, 1998; Schoenwald, Sheidow, & Letourneau, 2004). In addition, the utility of the TBRS-C is limited in one important respect: It does not currently possess a “red-line” score that serves as a benchmark for determining whether a given session was faithful to the treatment model. Such benchmarks can be valuable for training purposes in both research and practice contexts (Dobson & Shaw, 1993), but developing them requires a demanding process of consensus building among model experts and subsequent empirical verification that was beyond the scope of this study.

Study results support the utility of single global ratings of therapist competence in lieu of individual ratings of several therapeutic goals. The correlation between the global competence rating and the average competence rating across goals was medium to large in both CBT (r = .68) and MDFT (r = .79), indicating sizable overlap in the information captured by each method. Within each condition, the two methods produced an almost identical pattern of correlations with adherence and alliance variables. Other considerations favor the global rating method as well: It had acceptable interrater reliability in each condition, whereas many competence ratings for individual goals had poor reliability; it allowed judges to evaluate both the occurrence (what did

happen) and the nonoccurrence (what did not but should have happened) of therapeutic interventions; and it presented a lesser burden to code. All told, our recommendation for fidelity evaluation in field settings—and even in research settings when therapist competence is not the main focus—is the following: Assess adherence to the main therapeutic goals individually, but assess competence as one global dimension.

The next step in developing the TBRS-C is refining the item descriptions and scoring procedures for therapist competence to enhance their reliability. It is possible that interrater reliability would be stronger if the instrument were used with community therapists in routine clinical practice, who might be expected to show significantly greater variability in general therapy skills and in fidelity to manualized treatments. Another important step in examining the psychometric properties of the TBRS-C is predictive validity analyses: Do adherence and competence in treatment goals predict outcome? Previous studies on the same research sample showed that adherence to discrete intervention techniques was linked to adolescent outcomes in both treatments (Hogue, Dauber, Samuolis, & Liddle, 2006; Hogue, Liddle, Dauber, & Samuolis, 2004). The TBRS-C can also be used to track changes in model fidelity in the course of training a specific cohort of therapists (e.g., Crits-Christoph et al., 1998), replicating and adapting a given model in various clinical settings (Mowbray, Holter, Teague, & Bybee, 2003), or operationalizing therapy change processes in research on mechanisms of behavioral change (Doss, 2004). But clearly, the broadest potential value of the TBRS-C lies in adapting the instrument to measure adherence/competence for other evidence-based practices in everyday care for a wide variety of mental health disorders.

Acknowledgments

Preparation of this article was supported by grants R01 DA14571 (principal investigator: A. Hogue) and P50 DA07697 (principal investigator: H. Liddle) from the National Institute on Drug Abuse. The authors are extremely grateful to both teams of TBRS-C coders: for MDFT, Laura Alvarez-Cienfuegos, Mary Ekwall, Maria Dolores Fatas, Karly Gilbert, Oscar Ocasio, Lisette Sanabria-Velasquez, and Miguel Vilaro-Colon from the Roberto Clemente Family Guidance Center, along with Priscilla Chinchilla; for CBT, Jayme Albin, Peter Berzins, Roland Carstedt, Sue Manin, Leslie Sadoff, and Robert Udewitz from Behavioral Associates, along with Adam Fried. The authors thank Leyla Stambaugh, Leslie Alkalay, and Crystall Matthews for research contributions to this study, and Gayle Dakof and Cindy Rowe for providing feedback on the manuscript.

References

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

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.

Barber, J. P., Foltz, C., Crits-Christoph, P., & Chittams, J. (2004). Therapists' adherence and competence and treatment discrimination in the NIDA Collaborative Cocaine Treatment Study. *Journal of Clinical Psychology*, *60*, 29−41.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., et al. (2006). The role of therapist adherence, therapist competence, and alliance in perdicting outcome of individual drug counseling: Results from the National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*, *16*, 229−240.

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the Cognitive Therapy Adherence and Competence Scale. *Psychotherapy Research*, *13*, 205−221.

Carroll, K. M., Connors, G. J., Cooney, N. L., DiClemente, C. C., Donovan, D. M., Kadden, R. R., et al. (1998). Internal validity of Project MATCH treatments: Discriminability and integrity. *Journal of Consulting and Clinical Psychology*, *66*, 290−303.

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.

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

Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., et al. (1998). Training in cognitive, supportive–expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology*, *66*, 484−492.

Diamond, G. M., Liddle, H. A., Hogue, A., & Dakof, G. A. (1999). Alliance-building interventions with adolescents in family therapy: A process study. *Psychotherapy: Theory Research, Practice, & Training*, *36*, 355−368.

Diamond, G. S., & Diamond, G. M. (2002). Studying a matrix of change mechanisms: An agenda for family-based process research. In H. A. Liddle, D. A. Santisteban, R. F. Levant, & J. H. Bray (Eds.), *Family psychology: Science-based interventions* (pp. 41−46). Washington, DC: American Psychological Association: American Psychological Association.

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

Doss, B. D. (2004). Changing the way we study change in psychotherapy. *Clinical Psychology: Science and Practice*, *11*, 368−386.

Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling therapists from therapies. *Clinical Psychology: Science and Practice*, *6*, 10−32.

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

Garland, A. F., Hurlburt, M. S., & Hawley, K. M. (2006). Examining psychotherapy processes in a services research context. *Clinical Psychology: Science and Practice*, *13*, 30−46.

Hartley, D. E., & Strupp, H. H. (1983). The therapeutic alliance: Its relationship to outcome in brief psychotherapy. In J. Masling (Ed.), *Empirical studies of psychoanalytical theories*, *Vol. 1* Hillsdale, NJ: Analytical Press: Analytical Press.

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist–practitioner: Research and accountability in the age of managed care*. New York: Allyn & Bacon.

Henderson, C. E., Young, D. W., Jainchill, N., Hawke, J., Farkas, S., & Davis, R. M. (2007). Program use of effective drug abuse treatment

practices for juvenile offenders. *Journal of Substance Abuse Treatment*, *32*, 279−290.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. B., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, *65*, 821−833.

Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance abusing and dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research*, *1*, 171−184.

Hogue, A., Dauber, S., Faw, L., Cecero, J. J., & Liddle, H. A. (2006). Early therapeutic alliance and treatment outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *74*, 121−129.

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

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

Hogue, A., Liddle, H. A., Rowe, C., Turner, R. M., Dakof, G. A., & LaPann, K. (1998). Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. *Journal of Counseling Psychology*, *45*, 104−114.

Institute of Medicine. (2006). *Improving the quality of health care for mental and substance-use conditions*. Washington, DC: National Academic Press.

James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology*, *40*, 131−141.

Kazdin, A. E. (1999). Current (lack of) status of theory in child and adolescent psychotherapy research. *Journal of Clinical Child Psychology*, *28*, 533−543.

Levin, F. R., Owen, P., Stinchfield, R., Rabinowitz, E., & Pace, N. (1999). Use of standardized patients to evaluate the physicians in residence program: A substance abuse training approach. *Journal of Addictive Diseases*, *18*, 39−50.

Liddle, H. A. (2002a). Advances in family-based therapy for adolescent substance abuse. In L. S. Harris (Ed.), *Problems of drug dependence 2001* (pp. 113−115). Proceedings of the 63rd annual scientific meeting Bethesda, MD: National Institute on Drug Abuse (NIDA Monograph No. 182; NIH Publication No. 02-5097).

Liddle, H. A. (2002b). *Multidimensional family therapy for adolescent cannabis users. Cannabis youth treatment series*, *Vol. 5*. Rockville, MD: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration (DHHS Publication No. 02-3660).

Mark, T. L., Song, X., Vandivort, R., Duffy, S., Butler, J., Coffey, R., et al. (2006). Characterizing substance abuse programs that treat adolescents. *Journal of Substance Abuse Treatment*, *31*, 59−65.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, *72*, 1050−1062.

Morgenstern, J., Morgan, T. J., McCrady, B. S., Keller, D. S., & Carroll, K. M. (2001). Manual guided cognitive–behavioral therapy training: A promising method for disseminating empirically supported substance abuse treatments to the practice community. *Psychology of Addictive Behaviors*, *15*, 83−88.

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.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, *28*, 19−26.

Muck, R., Zempolich, K. A., Titus, J. C., Fishman, M., Godley, M. D., & Schwebel, R. (2001). An overview of the effectiveness of adolescent substance abuse treatment models. *Youth & Society*, *33*, 143−168.

Robbins, M. S., Turner, C. W., Alexander, J. F., & Perez, G. A. (2003). Alliance and dropout in family therapy for adolescents with behavior problems: Individual and systemic effects. *Journal of Family Psychology*, *17*, 534−544.

Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services*, *52*, 1190−1197.

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

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmstead, M., Vallis, T. M., Dobson, K. S., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology*, *67*, 837−846.

Shelef, K., Diamond, G. M., Diamond, G. S., & Liddle, H. A. (2005). Adolescent and parent alliance and treatment outcome in multidimensional family therapy. *Journal of Consulting and Clinical Psychology*, *73*, 689−698.

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

Startup, M., & Shapiro, D. A. (1993). Therapist treatment fidelity in prescriptive vs. exploratory psychotherapy. *British Journal of Clinical Psychology*, *32*, 443−456.

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, *5*, 439−458.

Turner, R. M. (1992). Launching cognitive–behavioral therapy for adolescent depression and drug abuse. In S. Budman, M. Hoyt, & S. Friedman (Eds.), *Casebook of brief therapy* (pp. 135−156). New York: Guilford: Guilford.

Waldron, H. B., & Kaminer, Y. (2004). On the learning curve: The emerging evidence supporting cognitive–behavioral therapies for adolescent substance abuse. *Addiction*, *99*(Suppl 2), 93−105.

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.

Williams, R. J., & Chang, S. Y., Addiction Centre Adolescent Research Group. (2000). A comprehensive and comparative review of adolescent substance abuse treatment outcome. *Clinical Psychology: Science and Practice*, *7*, 138−166.