Short form of the revised Vanderbilt Therapeutic Alliance Scale: Development, Reliability, and Validity

Karni Shelef, Gary M. Diamond

Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Online Publication Date: 01 July 2008

To cite this Article Shelef, Karni and Diamond, Gary M.(2008) 'Short form of the revised Vanderbilt Therapeutic Alliance Scale: Development, reliability, and validity', Psychotherapy Research, 18:4, 433 — 443

To link to this Article: DOI: 10.1080/10503300701810801

URL: http://dx.doi.org/10.1080/10503300701810801
Short form of the revised vanderbilt therapeutic alliance scale: Development, reliability, and validity

KARNI SHELEF & GARY M. DIAMOND

Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

(Received 7 August 2007; revised 4 November 2007; accepted 14 November 2007)

Abstract
This study presents the development and psychometric properties of a five-item version of the revised Vanderbilt Therapeutic Alliance Scale (VTAS-R Short Form). A total of 255 alliance ratings, derived from therapy sessions from 86 cases of family therapy for substance-abusing adolescents, were used to generate two separate subsamples. The first subsample was used to develop the five-item form and the second subsample to estimate the measure's reliability and concurrent and predictive validity. The VTAS-R Short Form evidenced high internal consistency, was highly correlated with the full-length VTAS-R, and predicted treatment completion and adolescents' days of cannabis use at 3 months follow-up. The findings suggest that the VTAS-R Short Form is a reliable and valid alternative to the full-length form.

Keywords: alliance; couples and family systems therapy; test development

A wealth of research conducted over the past 25 years attests to the robust association between the strength of the therapeutic alliance and treatment outcome in psychotherapy. The strength of the alliance, defined by Bordin (1979) as the degree to which the client and therapist agree on the goals and tasks of treatment and share a mutual, positive affective bond, has been shown to account for between 5% and 7% of the variance in treatment outcomes across a wide range of treatment approaches, clinical populations, and outcome parameters (Horvath, 1994, 2000; Horvath & Luborsky, 1993; Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000; Meier, Barrowclough, & Donnall, 2005), with alliance measured early in therapy being a slightly better predictor of outcome than alliance measured at mid-therapy or averaged across treatment (Horvath, 1994, 2001; Horvath & Symonds, 1991). Because of its consistent, although modest, ability to predict treatment outcome, the therapeutic alliance has become the most studied process variable in psychotherapy research.

A number of alliance measures have been developed to assess the alliance from various perspectives, including client self-report, therapist report, and observer ratings. In most studies, client self-report of the alliance has been shown to be most predictive of treatment outcome followed by observer ratings, with therapist ratings showing the least predictive validity (Horvath, 2000; Horvath & Symonds, 1991). However, among certain clinical populations, observer ratings may be most valid. For example, prior research on both adolescents (Kendall, 1994; Shelef, Diamond, Diamond, & Liddle, 2005) and substance-abusing adults (Fenton, Cecero, Nich, Frankforter, & Carroll, 2001) indicates that these groups report unusually high ratings of the alliance. It may be that, for various reasons, such clients are particularly susceptible to demand characteristics. In any event, uniformly high scores can lead to a ceiling effect and low variance, resulting in lower correlations between alliance and outcome. Indeed, prior studies with adolescents and substance-abusing adults have shown that observer ratings of the alliance were both more normally distributed and more predictive of outcome than client self-report, suggesting that observer ratings may be the measurement procedure of choice in studies with adolescents and substance abusers (Fenton et al., 2001; Shelef et al., 2005). Observer ratings of the alliance may also be preferred by psychotherapy process researchers interested in within-session changes. Such research requires multiple measurements of the alliance during the course of a given therapy session. Repeatedly interrupting the sessions to obtain the client's self-report would hopelessly interfere with...
the therapeutic process. Using observers to rate the alliance is intrusive and diminishes the threat of measurement by treatment effects.

One commonly used observer-based alliance measure is the Vanderbilt Therapeutic Alliance Scale (VTAS; Hartley & Strupp, 1983). The original VTAS consists of 44 items comprising three theoretically based subscales: (a) Therapist Contribution (18 items); (b) Patient Contribution (14 items); and (c) Therapist–Patient Interaction (12 items). The content of the items was drawn from several conceptual frameworks. Items were based on Bordin’s (1979) definition of the alliance as composed of bonds, goals, and tasks as well as on Greenson’s (1965) view of the alliance as a product of clients’ motivation to overcome their problems combined with a sense of helplessness. Additional contributions to item content came from Lang’s (1973) emphasis on the degree of openness and vividness with which clients report their experiences and Luborsky’s (1976) emphasis on the importance of clients and therapists developing a common framework to understand the possible causes and potential alleviators of clients’ problems. In their study examining VTAS ratings from 124 therapy sessions of 28 male college students suffering from depression, anxiety, and social difficulties, Hartley and Strupp (1983) found that the scale demonstrated adequate interrater reliability and agreement, based on product-moment correlations between the judges, interclass correlations, and percentage of exact and one-step discrepancy agreements. Internal consistency was reported to be high, as indicated by a coefficient alpha of .95, and the scale also demonstrated adequate predictive validity, as found in the association between changes in early treatment alliance and outcome. A factor analysis of the scale conducted by its developers did not confirm the three a priori theoretical subscales’ structure. Instead, six principal components were identified, four of which reflected patient contribution items (i.e., patient resistance, patient motivation, patient responsibility, and patient anxiety) and two reflected therapist contribution items (i.e., positive therapeutic climate, therapist intrusiveness). The therapist–patient interaction items contributed to almost all of the components (Hartley & Strupp, 1983). Later studies revealed a simpler two-factor structure of the VTAS: patient contribution (including patient and patient–therapist interaction items) and therapist contribution (Krupnick et al., 1996).

Frieswyk et al. (1986) have noted the conceptual distinction between therapist interventions designed to promote the alliance, and the strength of the alliance itself, as reflected by client behavior. Consequently, G. M. Diamond, Liddle, Dakof, and Hogue (1996) revised the VTAS by eliminating the items from the Therapist Contribution subscale and including only the 26 items from the Patient Contribution and Therapist–Patient Interaction subscales. The VTAS-R has been implemented in a number of studies (G. M. Diamond, Liddle, Hogue, & Dakof, 1999; Meyer et al., 2002; Robbins, Turner, Alexander, & Perez, 2003; Shelef et al., 2005) and has demonstrated adequate reliability and construct and predictive validity. Factor analyses of the VTAS-R suggest that it represents one general alliance dimension.

One disadvantage of the VTAS-R and other observer-rating alliance scales is that they require a tremendous amount of resources. Raters must be recruited and trained, a process that can last months. Once trained, multiple raters typically code each study session. A recording of an hour-long session can take a rater 2 hr or more to code because the coding process involves watching the entire session at least once and then making a determination regarding the score for each item. Understandably, alliance studies implementing observer-based measures often take a year or more to conduct and involve considerable costs. Factors such as these may discourage researchers from implementing observer-based measures, in spite of the empirical evidence suggesting that observer ratings may be a more valid measurement procedure for certain clinical populations (Fenton et al., 2001; Shelef et al., 2005). For example, in a review of 24 studies on the role of the alliance in drug treatment, only three studies used observer measures to rate the alliance (Meier et al., 2005). One way to reduce such barriers is to develop a shorter version of the VTAS–R. A shorter version may decrease the amount of time and resources required to train raters and to actually code sessions.

The purpose of this study was to develop, examine the reliability of, and validate a five-item version of the VTAS-R. Participants were adolescent substance abusers and their parents who received multidimensional family therapy (MDFT; Liddle, 2002). In line with recommendations by Coste, Guillemot, Pouchot, and Fermanian (1997) for shortening existing composite measurement scales, and in an effort to overcome methodological limitations found in past research (Smith, McCarthy, & Anderson, 2000), we took both empirical and theoretical considerations into account when choosing which items to retain in the short form. Furthermore, we used two separate subsamples of alliance ratings: one to develop the short form and the other to examine its reliability and validity. More specifically, we used the first subsample of adolescent-therapist and parent-therapist alliance ratings to conduct a factor analysis of the VTAS-R and to select five items, based on both
theoretical and empirical considerations. Using a second subsample of adolescent–therapist and parent–therapist alliance ratings, we examined the internal consistency of the short form, the correlation between scores on the short- and full-length forms, and the degree to which short-form alliance scores predicted treatment completion and adolescents’ days of cannabis use at posttreatment and at 3 months follow-up.

Method

Participants

Clients. Clients were drawn from the Cannabis Youth Treatment (CYT) study (Dennis et al., 2002). The CYT study was funded by the Center for Substance Abuse Treatment and is the largest clinical trial for adolescent substance users conducted to date. During a 2-year period, 600 adolescents and their families were recruited and randomized to five different treatments administered at four different treatment sites across the United States. To be included in the CYT study, clients must have (a) been between the ages of 12 and 18 years; (b) self-reported one or more Diagnostic and Statistical Manual of Mental Health Disorders (4th ed.; DSM-IV; American Psychiatric Association, 1994) criteria for cannabis abuse or dependence; (c) used marijuana in the past 90 days (or 90 days before being in a controlled environment); (d) met American Society of Addiction Medicine (ASAM; 1996) patient placement criteria for Level I (outpatient) or Level II (intensive outpatient); and (e) had a primary caretaker willing to participate in the assessments and treatment if randomized to a family-based condition. Potential clients were excluded if they (a) had used alcohol 45 or more days of the 90 days before intake (or before being in a controlled environment); (b) had used other illicit drugs 13 or more of the 90 days before intake; (c) had an acute medical or psychological condition that prohibited full participation in treatment (e.g., severe depression or suicidal ideation); (d) appeared to have insufficient mental capacity to understand the consent form or participate in treatment; or (e) had a history of violent behavior or severe conduct disorder.

In the course of the CYT study, 100 adolescent substance abusers and their families received MDFT in one of two major U.S. cities: One was located on the East Coast and the other in the Midwest. Of the entire sample, 86 adolescents and their families were included in this study. Clients were primarily male \( n = 73; 85\% \); 42 (49%) were White, 40 (47%) were African American, and the remaining 4 (4%) were identified as “other”. Their average age was 16 years (range = 13–18). A total of 76 (88%) were enrolled in school, and 54 (63%) were under the supervision of the juvenile justice system. In addition, 43 (50%) were from single-parent families. Socioeconomic status was assessed using the percentage of poverty index developed by Hollingshead and Redlich (1958): 17 (20%) of the participants were classified as very poor, 17 (20%) as poor, 34 (39%) as working class, 10 (12%) as upper middle class, and 8 (9%) as upper class. A total of 55 (64%) evidenced clinical levels of externalizing symptoms, and 26 (30%) evidenced clinical levels of internalizing symptoms. The entire sample qualified for a DSM-IV diagnosis of either substance abuse or dependency based on self-report only \( n = 74 \) [86%] or a combination of self- and collateral report \( n = 12 \) [14%].

Therapists. Three therapists administered MDFT. All three had at least 5 years of previous clinical experience. Two of the therapists were master’s-level clinical social workers and one was a PhD-level clinical psychologist. Two were male and one was female. One male was African American, and the other two therapists were White. Their ages ranged from 35 to 40 years. All three clinicians received at least 1 year of supervised training in MDFT before treating study cases. Training sessions were videotaped and then reviewed by the clinical coordinator. Therapists were assigned cases only after they were certified as proficient in MDFT. Weekly supervision continued throughout the study and included the review of at least two therapy videotapes per month to prevent therapist drift. During videotape reviews, the clinical coordinators completed treatment specific rating forms to monitor adherence and provide feedback to therapists.

Alliance raters. Raters were 10 undergraduate psychology students, ranging in age from 23 to 26 years. The group consisted of three males and seven females.

Treatment

MDFT is a 12-week, family-based, multisystemic approach to treating substance-abusing adolescents and their families. It is a carefully constructed, manualized, empirically based clinical model (Liddle, 2002; Liddle, Dakof, & Diamond, 1991; Liddle & Diamond, 1991). Its theoretical roots are in the integrative structural–strategic family therapy tradition (Fraser, 1982; Stanton, 1981; Todd, 1986). However, MDFT also exists within the ecological
tradition and, therefore, addresses the multiple subsystems in which people reside (e.g., individual, marital, parental, extrafamilial, peer, sibling), both within and beyond the context of the family. The overall goals for treatment are based on empirical research on normative adolescent development and developmental psychopathology. Adolescent substance abuse is understood as existing in a context of other, interrelated problems, such as poor relationships, deficits in cognitive and problem-solving skills, learning and school difficulties, low self-esteem, family stress or dysfunction, and movement onto a trajectory of failure and incompetence. In accordance with research findings on parenting and family environments associated with adolescent drug use (Baumrind, 1991), MDFT focuses on issues such as interdependence (Steinberg, 1990) and autonomy-connectedness (Grotevant & Cooper, 1983) in the parent-adolescent relationship. Moreover, all problem definitions and respective interventions and change processes are conceived of as multifaceted (e.g., cognitive, affective, behavioral).

The specific goals of treatment are derived from each family’s unique presentation. However, in every case, therapists are instructed to spend the early part of treatment developing a strong and separate working alliance with each of the family members (e.g., adolescent, parent) as well as with extrafamilial systems (Liddle & Diamond, 1991). This is accomplished through conjoint family sessions as well as individual sessions with the adolescent or parents.

These alliances are meant to engage family members in the therapy process, instill a sense of purpose and competence, generate hope and collaboration, and build trust and confidence in the therapist. Once the alliance has been sufficiently formed and the goals and tasks of the treatment identified, the therapist uses these alliances as leverage in challenging family members’ attributions and behaviors. In some instances, the therapist may suggest that parents adapt or revise certain parenting practices (Schmidt, Liddle, & Dakof, 1996). In other instances, the therapist may urge parents and adolescents to disclose vulnerable feelings to each other in the context of reconnection episodes (i.e., in-session enactments) designed to increase positive parent-adolescent interaction (G. S. Diamond & Liddle, 1996, 1999; Liddle, Rowe, Dakof, & Lyke, 1998).

Several randomized clinical and prevention trials have established the efficacy of MDFT with moderate to severe drug-abusing adolescents (U.S. Department of Health and Human Services [USHHS], 2002; Waldron, 1997; Weinberg, Rahdert, Colliver, & Glantz, 1998). These studies indicate that MDFT is superior or equal to other types of well-established drug abuse treatments (e.g., cognitive–behavioral therapy; adolescent group therapy) at improving a number of target areas (e.g., drug use, problem behaviors, family functioning) by therapy termination and at follow-up (Dennis et al., 2004; Liddle et al., 2001; USHHS, 2002). In addition, findings from the CYT study show that MDFT was as effective as other standardized drug treatments at reducing adolescents’ drug use and drug-related symptoms (Dennis et al., 2004).

Measures

Adolescent-therapist and parent-therapist observed alliance. Observer ratings of the two alliances were assessed using the VTAS-R (G. M. Diamond et al., 1996). As mentioned earlier, the VTAS-R includes only the 26 items covering the Patient Contribution (e.g., “To what extent did the patient acknowledge that he had a problem which the therapist could help him with?”) and the Therapist–Patient Interaction (e.g., “To what extent did the therapist and patient together share a common viewpoint about the definition, possible causes, and potential alleviation of the patient’s problems?”) subscales. Nine of the items are reversed, and all are scored on a 6-point Likert-type scale, ranging from 0 (not at all) to 5 (a great deal).

The items pertain to all three aspects of the therapeutic alliance (i.e., bond, goals, and tasks), although prior factor analyses of the VTAS and other alliance instruments have failed to distinguish among the bond, goals, and tasks dimensions (Andrusyna, Tang, DeRubeis, & Luborsky, 2001; Hatcher & Barends, 1996; Krupnick et al., 1996).

As a consequence of training raters for various previous studies, a number of minor revisions were made to the VTAS-R to clarify items or improve interrater reliability. First, we eliminated one item—“To what extent did the patient make an effort to carry out therapeutic procedures suggested by the therapist?”—because carrying out therapeutic procedures is considered an indication of agreement on tasks, and a separate, more explicit item pertaining to agreement on tasks existed within the scale. Also, two additional items —“To what extent did the therapist and patient accept their different roles and responsibilities as part of their relationship?” and “To what extent did the patient expect the therapist to change him without accepting his own responsibility for the therapy?”—were eliminated because expert trainers had difficulty clarifying the meaning of these items and, consequently, coders had difficulty in coding them reliably. As a result, the final version of the VTAS-R implemented in this study included 23 items. Internal consistency was very high, as indicated by Cronbach’s coefficient alpha of
.95. Individual-item interclass correlation coefficients (ICCs; Shrout & Fleiss, 1979) ranged from .55 (i.e., “To what extent did the patient show a willingness to explore his/her own contribution to his/her life situation?”) to .87 (i.e., “To what extent did the patient seem to identify with the therapist’s method of working, so that he/she assumed part of the therapeutic task?”). According to guidelines provided by Shrout and Fleiss, the interrater reliability estimate was excellent for 14 of the items, good for eight items, and fair for one.

**Substance use and symptomatology.** Participant characteristics, substance use, and symptomatology were measured with the Global Appraisal of Individual Needs (GAIN; Dennis, 1999). The GAIN is a standardized clinical assessment battery covering eight main domains (background, substance use, physical health, risk behaviors, mental health, environment, legal, and vocational) and has been normed on both adults and adolescents (Dennis, Scott, Godley, & Funk, 1999, 2000). For the purposes of this study, three variables were extracted from the GAIN: days of cannabis use, internalizing problems scale scores, and externalizing problems scale scores.

**Days of cannabis use.** Days of cannabis use refers to the number of days the adolescent used marijuana or hashish in the past 90 days regardless of frequency or amount of use per day. Adolescents’ self-report of days of cannabis use showed high test–retest reliability \( r = .74 \) (Dennis et al., 2002) and was consistent with family/collateral reports, on-site urine tests, and gas chromatography/mass spectrometry (GC/MS) tests for delta-9-tetrahydrocannabinol (\( \Delta ^{9} \)-THC) at intake and various follow-up waves \( \kappa = .70-.90 \); Buchan, Dennis, Tims, & Diamond, 2002).

**Internalizing problems scale.** This scale is a count of diagnoses of internalizing disorders in the past year, including general anxiety disorder (diagnosis is based on nine items), major depressive disorder (diagnosis is based on 16 items), traumatic stress disorder (diagnosis is based on 13 items), and suicide thoughts (diagnosis is based on four items). For the purpose of our analysis, the scale was recoded into a dichotomous variable (no internalizing diagnosis vs. one or more internalizing diagnoses).

**Externalizing problems scale.** This scale is a count of diagnoses of externalizing disorders in the past year, including attention-deficit/hyperactivity disorder (diagnosis is based on 18 items) and conduct disorder (diagnosis is based on 15 items). For the purpose of our analysis, the scale was recoded into a dichotomous variable (no externalizing disorder vs. one or more externalizing disorders).

**Procedure**

**Session sampling criteria.** First and late sessions (i.e., Session 11 or higher) were excluded from the analyses because of the unique themes and processes characteristic of such sessions. For example, first sessions in MDFT are typically devoted mainly to introductions and socialization to treatment, whereas the focus in late (or last) sessions is typically on the progress achieved through treatment and farewells. Ratings of the alliance in such sessions may be influenced by their unique content or process; consequently, certain VTAS-R items become less relevant.

**Samples.** Two subsamples of alliance ratings were used to carry out the two stages of VTAS-R Short Form development: (a) item analysis and selection and (b) estimating reliability, concurrent validity, and predictive validity. The two subsamples of alliance ratings were drawn from a pool of 433 available videotaped sessions from 86 cases receiving MDFT in the CYT study. For each of the 86 cases, there were on average five videotaped sessions available. One adolescent–therapist and one parent–therapist alliance rating from each case was included in each of the two subsamples. The sessions from which the alliance ratings were derived were chosen randomly from all videotaped sessions available for that given case. However, care was taken so that the alliance ratings included in Subsample 1 were derived from different sessions than the alliance ratings included in Subsample 2.

Subsample 1, which was used to develop the short form of the VTAS-R, included 127 alliance ratings. Of these ratings, 69 were of the adolescent–therapist alliance and 58 were of the parent–therapist alliance. Of the adolescent alliance scores, 80% were coded in individual sessions (i.e., only the adolescent was present in the session), and the remaining 20% were coded in conjoint sessions (i.e., both the adolescent and at least one of the parents were present). For parent alliance scores, 67% were coded in individual sessions and 33% were coded in conjoint sessions. Subsample 2, which was used to examine the reliability, concurrent validity, and predictive validity of the short form, included 128 alliance ratings. Of these ratings, 70 were of the adolescent–therapist alliance and 58 were of the parent–therapist alliance. Of the adolescent alliance ratings, 80% were coded in individual sessions, and the remaining 20% were...
coded in conjoint sessions. Of the parent alliance ratings, 55% were coded in individual sessions, and the remaining 45% were coded in conjoint sessions.

Alliance rater training. Rater training was conducted separately for those raters rating the adolescent–therapist alliance and for those rating the parent–therapist alliance. Training sessions were 4 hr in duration and were conducted three times a week over the course of 1 month. Raters were first given the manual to study. After becoming familiar with the manual, they were then shown and asked to score dozens of videotaped instances drawn from actual sessions representing various levels of alliance strength. Training tapes were not included in the study sample. Ongoing discussions were conducted to clarify scoring dilemmas. After 1 month of training, raters were given five practice sessions to rate. Analyses showed that for these practice sessions, raters achieved very good to excellent inter-rater reliability, ICC2,2 > .85, for both alliances and, therefore, were deemed competent to begin coding actual study tapes.

Alliance coding procedure. Adolescent–therapist and parent–therapist alliances were coded independently, with two raters separately coding each alliance for each session. When rating the alliance in conjoint sessions, coders were instructed to focus only on the adolescent or the parent and on his or her interaction with the therapist. Alliance scores were generated for one parent in each family. In 68% of the sessions, only one parent was present, and in 83% of these instances, it was the mother. In the remaining 32% of cases, in which both parents were present, alliance scores were based on the behavior of the more outspoken of the two parents, which was the mother 71% of the time.

Final alliance scores for each session were calculated by averaging the scores of the two raters. Raters were assigned videotapes in rotating, random pairs. Weekly recalibration sessions were provided to avoid rater drift. Raters were naive to the session number being coded and to the purpose and hypotheses of the study.

Results

The procedure for generating the VTAS-R Short Form and testing its psychometric properties involved two phases: (a) development of the short form, including factor analysis, item analysis, and item selection; and (b) establishing reliability estimates for the short form, concurrent validity between the short- and full-length forms of the scale, and examining the predictive validity of the short form.

Preliminary Analyses

Session number. Both subsamples were composed of ratings from Sessions 2 to 10. To examine whether Subsamples 1 and 2 evidenced the same distribution of session number, we conducted a chi-square analysis. Results indicated that there was no difference between the subsamples on this dimension, \( \chi^2(8, N = 255) = 4.89, ns \).

Session constellation. Because MDFT is a family-based treatment, sessions could include the adolescent or parents only (i.e., individual sessions) or both the adolescent and parents together (i.e., conjoint sessions). To examine whether the frequency of individual and conjoint sessions differed among the two subsamples, a chi-square analysis was conducted. Results indicated that the two subsamples did not differ in terms of session constellation, \( \chi^2(1, N = 255) = 0.87, ns \).

Adolescent alliance by session constellation. Independent-samples \( t \) tests were conducted to examine whether the mean alliance score was different when assessed in individual versus conjoint sessions. Results indicated that there was no difference in the mean alliance score of these sessions for adolescent alliance, \( t(137) = 0.47, ns \), or for parent alliance, \( t(114) = 0.05, ns \).

Adolescent alliance by gender. Independent-samples \( t \) tests were conducted to examine whether there was a difference between male and female adolescent alliance scores. Results indicated that adolescent alliance scores did not differ significantly according to gender, \( t(67) = 0.13, ns \).

Adolescent alliance by race. Independent-samples \( t \) tests were conducted to examine whether alliance scores were different for Caucasian versus African American adolescents. Results indicated that alliance scores for Caucasian adolescents (\( M = 3.48, SD = .67 \)) were significantly higher than for African American adolescents (\( M = 2.86, SD = .96 \)), \( t(65) = -3.07, p < .01 \). Consequently, to control for the association between race and the therapist-adolescent alliance, we included race in the main analyses as a covariate.

Adolescent alliance by family structure. Independent-samples \( t \) tests were conducted to examine whether alliance scores were different for adolescents with one parent living in the home (i.e., single-parent
families) versus those in a two-parent home. Results indicated that adolescent alliance scores did not differ significantly according to family structure, \( t(67) = 0.81, \) not.

**Adolescent alliance by diagnosis.** To examine whether adolescent alliance scores were associated with adolescents’ diagnoses, we conducted a one-way analysis of variance. The three groups compared were adolescents with externalizing diagnoses only, those with co-occurring externalizing and internalizing diagnoses, and those with neither externalizing nor internalizing diagnoses. Only three adolescents had internalizing diagnoses with no co-occurring externalizing diagnoses and were, therefore, not included in the analysis. Results indicated that adolescent alliance scores did not differ significantly according to type of diagnosis, \( F(2, 120) = 0.28, \) not.

**Parent alliance.** Independent-samples \( t \) tests indicated that there were no differences in the mean parent alliance score for those sessions in which only one parent was present versus those sessions including two parents, \( t(114) = -0.92, \) not, nor were there any differences between mothers’ versus fathers’ mean alliance scores, \( t(114) = 1.08, \) not.

**Development of the VTAS-R Short Form**

**Elimination of items.** To choose which items to include on the short form, we undertook a number of steps. First, we decided to remove one item—“To what extent did the patient express feeling better since beginning of therapy?”—because it was deemed to measure treatment progress rather than the alliance per se.

Next, we eliminated items with the most highly skewed (\( >1.5 \)) distributions of scores. These items received the lowest rating of 0 at least 75% of the time. In accordance with these criteria, the following six items were omitted: “To what extent did the patient act in a hostile, attacking or critical manner toward the therapist?”; “To what extent did the patient become so anxious in the session that it interfered with the therapeutic task?”; “To what extent did the patient show evidence that she would rather be somewhere else?”; “To what extent did the therapist and patient together seem to be engaged in a power struggle?”; “To what extent did the patient express directly or indirectly the possibility of premature termination?”; and “To what extent did the therapist and patient together have awkward silences or pauses in their conversation?”

**Factor analysis.** To explore the factor structure of the remaining 16 items, a principal-components analysis was conducted. The analysis yielded a one-factor solution (eigenvalue = 10.23), accounting for 63.94% of the variance. One item (“To what extent did the therapist and patient refer back to experiences they have been through together”) did not load on this factor (\( h^2 < .1 \)) and was, therefore, eliminated. All of the remaining items had loadings above .65.

**Final Item Selection**

Our a priori goal was to construct a five-item scale. Because the remaining 15 items were highly intercorrelated and represented a single, reliable, face-valid factor, we presumed that any five-item combination would account for approximately the same amount of variance in the full-length form alliance score. Consequently, we based our choice of five items not only on statistical criteria but also on conceptual considerations. More specifically, our first step was to choose five items whose face validity best reflected Bordin’s three theoretical constructs (i.e., bonds, goals, and tasks) and that reflected both the patient and the therapist-patient interaction contributions to the alliance, two theoretical orientations that guided the development of the original VTAS (Hartley & Strupp, 1983). We also chose items that required the least inference to code. The five items chosen are presented in Table I.

Once we made our initial selections, we sought to confirm that each of the five items evidenced sufficient reliability at the item level. Interrater reliability ranged from good to excellent, as indicated by average ICCs ranging from .72 to .87 (see Table II). Next, to preserve a high level of internal consistency in the short form, we examined all five items to ensure that the corrected item-total correlation between each of the items and the full-length scale score (i.e., total score excluding the correlated item) was sufficiently high. Item-total correlations were both high (\( r > .70 \)) and significant (see Table II).

<table>
<thead>
<tr>
<th>Table I. Vanderbilt Therapeutic Alliance Scale–Revised Short Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extend did the patient:</td>
</tr>
<tr>
<td>1. Indicate that she experiences the therapist as understanding and supporting her. (B)</td>
</tr>
<tr>
<td>2. Seem to identify with the therapist’s method of working, so that she assumes part of the therapeutic task. (T)</td>
</tr>
<tr>
<td>3. Act in a mistrustful or defensive manner toward the therapist. (B)</td>
</tr>
<tr>
<td>4. Share a common viewpoint about the definition, possible causes, and potential alleviation of the patient’s problems. (G)</td>
</tr>
<tr>
<td>5. Agree upon the goals and/or tasks for the session. (G/T)</td>
</tr>
</tbody>
</table>

Note. B = bond; G = goals; T = tasks.
Reliability and Validity

To examine the reliability and concurrent and predictive validity of the five-item form, we used alliance ratings from Subsample 2 (see Samples subsection). Because our goal was to establish reliability and validity estimates for both adolescent and parent populations, all subsequent analyses were conducted separately for adolescent and parent alliance scores.

*Internal consistency.* Internal consistency for the VTAS-R Short Form was high, as indicated by a coefficient alpha of .90 for adolescents and .91 for parents.

*Concurrent validity.* To examine the concurrent validity of the VTAS-R Short Form, we examined the overlapping variance between the short- and full-length form scores. We chose not to run a simple correlation between the two scores because that would have led to an overestimation of the shared variance: Any error or random variance associated with the five items of the short form would, in fact, be included in both scores (Smith et al., 2000). To avoid such inflated estimates, we opted to correlate the five-item score with the score of the 18 remaining items (i.e., the full-length scale score, excluding the five items of the short form). The correlation between the five-item and 18-item scores was extremely high for both adolescents \(r = .94, p < .001, n = 70\) and parents \(r = .90, p < .001, n = 58\).

Predictive Validity

To examine whether the short form evidenced adequate predictive validity, we examined the association between short-form alliance scores and treatment outcome. Because each of the therapists in our study administered treatment to multiple families, it was plausible that ratings of sessions from the same therapist would share some degree of variance. Consequently, in the following analyses, we account for the nested nature of our data.

*Alliance and treatment completion.* To examine whether the adolescent alliance or parent alliance predicted treatment completion, we conducted a hierarchical binary logistic regression analysis with treatment completion as the dependent variable. To account for the nested nature of the data, therapist was entered as a categorical covariate in the first block using two dummy variables. The second block included adolescent alliance and parent alliance as predictors. Treatment completion (i.e., the dependent variable) was defined as completion of seven sessions or more (for rationale see Shelef et al., 2005). Because the dropout criterion was set at seven sessions, only alliance scores derived from Sessions 2 to 7 were included in this analysis. In addition, because both adolescent and parent alliances were entered as predictors in this model, we included only cases for which observer ratings of both alliances were available. Consequently, the subsample for this analysis included 34 cases, which were composed of 28 treatment completers and six cases that terminated treatment prematurely. Results indicated that the full regression model had a pseudo \(R^2\) of .39 and approached significance, \(\chi^2(4, N = 34) = 9.27, p = .06\). Parent alliance was marginally significant in predicting treatment completion, \(\text{Exp}(B) = 2.34, \text{Wald} Z = 3.02, p = .08\), so that for each one-unit increase in the parent alliance score, the family was 2.34 times more likely to complete treatment. Adolescent alliance did not predict treatment completion.

*Race and treatment completion.* Our preliminary analysis yielded a significant association between the therapist–adolescent alliance and race. This finding suggests that race may have an impact on the association between alliance and treatment completion. Unfortunately, because of the size of our sample and the distributions of race and completer status (i.e., there were only six dropout cases involving four African American and two Caucasian families), we could not include race as an additional predictor in the logistic regression analysis described previously. However, in an effort to examine the direct association between race and treatment completion, we conducted a chi-square analysis. Results indicated that the association between race and treatment completion was not significant, \(\chi^2(1, N = 31) = .30, \text{ns}\).

*Alliance and outcome.* To examine the effect of adolescent alliance on treatment outcome (i.e., days of cannabis use) at posttreatment and at 3 months

<table>
<thead>
<tr>
<th>Variable</th>
<th>Average ICC(_{2,2})</th>
<th>Item-total correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Understanding and supporting</td>
<td>0.85*</td>
<td>0.87*</td>
</tr>
<tr>
<td>2. Identify with method</td>
<td>0.87*</td>
<td>0.91*</td>
</tr>
<tr>
<td>3. Mistrustful or defensive</td>
<td>0.73*</td>
<td>0.82*</td>
</tr>
<tr>
<td>4. Common viewpoint</td>
<td>0.81*</td>
<td>0.86*</td>
</tr>
<tr>
<td>5. Goals and/or tasks</td>
<td>0.72*</td>
<td>0.72*</td>
</tr>
</tbody>
</table>

Note. ICC = interclass correlation coefficient. \(^*p < .001\).
follow-up, two multilevel models were computed using mixed-model analysis. Only treatment completers were included in this analysis \((n = 56)\). The dependent variables were days of cannabis use at posttreatment and at 3 months follow-up in the first and second models, respectively. In both models, pretreatment days of cannabis use and race were entered as covariates, adolescent alliance was entered as the fixed effect, and both were nested within therapist. A maximum likelihood estimation procedure was used, an unstructured covariance structure was adopted, and degrees of freedom were calculated using the between-within method. In the model examining days of cannabis use at 3 months follow-up, results indicated that after adjusting for therapist effect and controlling for pretreatment levels of cannabis use and race, the effect for adolescent alliance was significant, with an estimated slope of \(-7.64, F(1, 50) = 6.76, p < .05\). As expected, higher levels of adolescent–therapist alliance were associated with lower levels of cannabis use at 3 months follow-up. In the model examining days of cannabis use immediately posttreatment, only the effect for pretreatment levels of cannabis use was significant, with an estimated slope of 0.23, \(F(1, 54) = 5.69, p < .05\). Adolescent alliance did not have a significant effect on posttreatment levels of cannabis use.

**Discussion**

A vast amount of research shows that the therapeutic alliance measured early in therapy predicts treatment outcome across a wide range of therapeutic models and clinical populations. Although in most cases it is clients’ reports of the alliance that are most strongly associated with outcome, for some populations, such as other-referral substance abusers and adolescents, observer-rated alliance may be most valid. One of the most commonly used instruments for generating observer ratings of the alliance has been the VTAS. However, the original VTAS includes 44 items and the VTAS-R includes 26 items, requiring substantial resources to both train raters and to code sessions. This study was designed to create a reliable and valid five-item short form of the VTAS-R that reflected both Bordin’s (1979) three dimensions of the alliance (i.e., goals, tasks, and bond), and the patient’s and therapist–patient interaction’s contributions to the alliance (Hartley & Strupp, 1983).

Results indicated that the VTAS-R Short Form was highly reliable, was highly correlated with the full-length form, and predicted treatment completion and adolescents’ subsequent cannabis use and substance abuse and dependency symptoms. More specifically, the five-item short form was highly correlated with the full-length form for both adolescents and parent alliance scores \((r = .94\) and .90, respectively). These correlations are extremely high, and they equal or surpass the correlations found between most short and long forms of instruments designed to measure a variety of psychological constructs, such as the Brief Symptom Inventory (Derogatis & Savitz, 2000), a revised short version of the Working Alliance Inventory (WAI; Hatcher & Gillaspy, 2006), and a short form of the Wechsler Adult Intelligence Scale-Revised (Crawford, Mychalkiw, Johnson, & Moore, 1996). In addition, parent alliance scores derived from the short form were predictive of treatment completion, and adolescent alliance scores derived from the short form predicted days of cannabis use at 3 months follow-up, indicating that the short form demonstrated predictive validity for both adolescents and parents.

These findings were consistent with results from a previous study on the same sample, in which full-length VTAS-R parent alliance scores predicted treatment completion and full-length VTAS-R adolescent alliance scores predicted adolescents’ days of cannabis use at 3 months follow-up but not at posttreatment (Shelef et al., 2005). The consistency of the findings across the two studies is remarkable in light of the fact that two different, mutually exclusive subsamples of sessions were analyzed in the two studies.

Together, these findings suggest that the five-item short form may be interchangeable with the full-length form of the VTAS-R. Such findings are consistent with findings from other studies developing short versions of existing, psychometrically sound alliance measures, such as the WAI (Busseri & Tyler, 2003; Tracey & Kokotovic, 1989).

A number of methodological strengths increase our confidence in the findings reported. First, we used both empirical and conceptual considerations in the scale development process. Consequently, the VTAS-R Short Form is not only reliable and valid but clearly reflects central constructs of the alliance also found in other commonly used alliance measures. Second, we used two separate subsamples to conduct the two stages of the scale development/evaluation process, thus increasing the external validity of the findings. Finally, our analyses were conducted on a large number of alliance ratings, including both adolescents and parents, suggesting that the instrument is useful for both adolescent and adult populations.

Nevertheless, certain aspects of the study limit our interpretation of the results. First, the short-form scores in this study were derived from full-length VTAS-R ratings. The fact that the two forms were not administered independently hinders our ability...
to interpret the correlation between them. Although care was taken to avoid overestimating the correlation by excluding the five items of the short form from the full-scale score, there may still be systematic error effects (e.g., from neighboring items) that are influencing both scores. Further research is needed to examine the reliability and validity of the short form when it is administered in its five-item format. Second, the two subsamples used in the study were not independent. Although the two subsamples consisted of alliance ratings from different sessions, the sessions were derived from the same group of clients. Third, empirical data are needed to estimate the amount of time required to train raters to use the short form and the amount of time required to use the short form to rate sessions. Such data will allow us to calculate the net benefit in terms of training and coding time. Fourth, the psychometrics of the short form were examined in the context of one specific type of therapy delivered to a specific population: adolescent substance abusers and their parents. To examine the generalizability of the findings, more research is needed to examine the short form’s reliability and validity in the context of less structured or different treatment modalities and different clinical populations.

Even in light of these limitations, the findings from this study suggest that the VTAS-R Short Form is an excellent alternative to the full-length VTAS-R. Although implementing this five-item observer-based alliance measure still requires training raters and having raters view videotaped or live therapy sessions, the small number of items promises to reduce the time it takes to rate sessions and, ultimately, the human resources and financial cost of conducting such studies. Reducing the burden and cost of using the VTAS-R may encourage investigators to collect observer ratings of the therapeutic alliance, so critical to psychotherapy process research with other-referred clinical populations.

Acknowledgements

This study was supported by Israel Science Foundation Grant 865/01 awarded to the second author.

References


