



Published in final edited form as:

*Psychol Assess.* 2013 March ; 25(1): 146–156. doi:10.1037/a0029715.

## A Review of Treatment Adherence Measurement Methods

Sonja K. Schoenwald and

Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina

Ann F. Garland

University of California, San Diego and Child and Adolescent Services Research Center, Rady Children's Hospital

### Abstract

Fidelity measurement is critical for testing the effectiveness and implementation in practice of psychosocial interventions. Adherence is a critical component of fidelity. The purposes of this review were to catalogue adherence measurement methods and assess existing evidence for the valid and reliable use of scores they generate and feasibility of use in routine care settings.

**Method**—A systematic literature search identified articles published between 1980–2008 reporting studies of evidence-based psychosocial treatments for child or adult mental health problems, and including mention of adherence or fidelity assessment. Coders abstracted data on the measurement methods and clinical contexts of their use.

**Results**—341 articles were reviewed in which 249 unique adherence measurement methods were identified. These methods assessed many treatment models, although more than half (59%) assessed Cognitive Behavioral Treatments. The measurement methods were used in studies with diverse clientele and clinicians. The majority (71.5%) of methods were observational. Information about psychometric properties was reported for 35% of the measurement methods, but adherence-outcomes relationships were reported for only ten percent. Approximately one third of the measures were used in community-based settings.

**Conclusions**—Many adherence measurement methods have been used in treatment research; however, little reliability and validity evidence exists for the use of these methods. That some methods were used in routine care settings suggests the feasibility of their use in practice; however, information about the operational details of measurement, scoring, and reporting is sorely needed to inform and evaluate strategies to embed fidelity measurement in implementation support and monitoring systems.

### Keywords

Treatment adherence; measurement methods; fidelity; implementation

---

Making effective mental health treatment more widely available in routine practice is a public health priority. Research on the ingredients and processes necessary and sufficient to increase the adoption, adequate implementation, and sustainability of even one evidence-based treatment, much less the diversity of treatments subsumed under the evidence-based moniker, is in its early stages. Conceptual models and heuristic frameworks based on pertinent theory and research in other fields suggest recipes for dissemination, implementation, and sustainability of effective treatment will require attention to the

interplay of ingredients at multiple levels of the practice and policy context. Scholars contributing to the emerging field of implementation science have identified the monitoring of treatment fidelity as among the features likely to characterize effective implementation support systems (Aarons, Hurlburt, & Horwitz, 2011).

In psychotherapy research, there are three components of treatment fidelity: therapist adherence, therapist competence, and treatment differentiation. Therapist adherence refers to the extent to which treatments as delivered include prescribed components and omit proscribed ones (Yeaton & Sechrest, 1981). Thus, the core task of adherence measurement is to answer the question “Did the therapy occur as intended?” (Hogue, Liddle, & Rowe, 1996, p.335). To facilitate and evaluate the larger scale adoption and implementation of evidence-based treatments in clinical practice, adherence measurement methods are needed that yield valid and reliable scores and can be incorporated with relatively low burden and expense into routine care. Elsewhere, we have taken the linguistic liberty of framing this issue as a need for measurement methods that are both effective (i.e., yield scores that can be used to make valid and reliable decisions about therapist adherence) and efficient (i.e., feasible to use in practice), and have identified attributes of both the measurement process and the clinical context pertinent to the development of such methods (Schoenwald, Garland, Chapman, Frazier, Sheidow, & Southam-Gerow, 2011a).

Over a decade ago, leading research and policy voices highlighted the need for development of more efficient fidelity (including adherence) measurement methods, suggesting that the lack of low-burden, inexpensive, fidelity indicators was a barrier to behavioral health care improvement (Hayes, 1998; Manderscheid, 1998). Likewise, leading psychosocial treatment researchers have highlighted the lack of evidence for effective (i.e. valid) measurement of treatment adherence, competence, and differentiation in randomized trials published in leading journals between 2000 and 2004 (Perepletchikova, Treat, & Kazdin, 2007).

Despite these limitations however, the dissemination and implementation of a variety of evidence-based psychosocial treatments is well underway, with and without adherence measurement methods that have previously been evaluated. Treatment adherence indicators are essential for stakeholders in mental health (clients, practitioners, payers) to determine whether client outcomes in routine care -- favorable or not -- are attributable to a particular treatment or treatment implementation failure. The extent to which adherence measurement methods have been developed, and are potentially effective and efficient for use with a range of clinical populations, treatment models, and treatment settings in routine care is not known.

The purposes of the current review were to catalogue extant adherence measurement methodologies for evidence-based psychosocial treatments and to characterize them with respect to their effectiveness (reliability and validity evidence) and efficiency (feasibility of using the methods in routine clinical care). Gaining an understanding of the range of available measurement methods and the extent to which existing measurement methods are characterized by both of these attributes, as well as the purposes of adherence measurement evolving in the context of dissemination and implementation efforts, represents a step in identifying gaps in the availability of sound and practical adherence measurement methods for evidence-based treatments needed to advance research on their dissemination and implementation (Schoenwald et al., 2011a).

## Method

### Search Strategies

Although the current study is not a meta-analysis, the methods used to identify the sample of articles about psychosocial treatments to be reviewed for evidence of adherence measurement were informed by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009) and the MARS (Meta-Analysis Reporting Standards: Information Recommended for Inclusion in Manuscripts Reporting Meta-Analyses; APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). Figure 1 presents the PRISMA flow diagram for the current review.

The population of articles from which the current sample was drawn was identified through searches of both the Medline and PsychInfo computerized databases. Initially, articles had to meet two criteria to be identified as eligible for inclusion in the review. (1) The English language articles were published in peer-reviewed journals between 1980 and 2008. Thus, excluded were dissertations, book chapters, and other unpublished work. (2) Articles reported on (a) use of an empirically supported psychosocial treatment for (b) mental health/behavioral health problems. With respect to (a), empirically supported psychosocial treatment models and programs were identified on the basis of published reviews (see, e.g., Bergin & Garfield, 1994; Chambless & Ollendick, 2001; Hibbs & Jensen, 1996; Kazdin & Weisz, 2003; Kendall & Chambless, 1998; Liddle, Santisteban, Levant, & Bray, 2002; Silverman & Hinshaw, 2008; Special Issue on Empirically Supported Psychosocial Interventions for Children, 1998; Stewart & Chambless, 2009; Weisz & Kazdin, 2010). With respect to (b), articles were excluded that reported on the use of the psychosocial treatment models and programs solely to treat medical conditions such as asthma, brain injury, diabetes, insomnia, smoking, or obesity.

The search terms used to identify in the computerized data bases a sample of articles meeting these two criteria, were (a) the name of each treatment model or program, or (b) the words, “adherence,” or “fidelity.” Articles were identified for possible inclusion if any of these terms appeared in the Title, Abstract, and Key Words. Using these search terms, 750 articles were identified for potential inclusion in the review. Screening of article content was conducted to eliminate articles reporting on treatment solely of health/medical conditions (selection criteria # 2b) and to identify articles that included information about how adherence or fidelity to treatment was assessed. This content screening process identified 397 articles as potentially eligible for the review. Of these, 41 were subsequently excluded because they reported on a therapy construct other than treatment adherence, such as therapist competence, patient adherence to a protocol, verbal interactions not related to adherence or fidelity, or verbal interaction styles not related to adherence or fidelity. Fifteen articles were identified as duplicates.

### Coding Procedures

**Manual**—The authors and a psychotherapy researcher with considerable expertise in psychosocial treatment literature meta-analysis coding developed the coding manual, entitled *Fidelity Measure Matrix Coding Manual* (Implementation Methods Research Group; IMRG, 2009). Variables were defined to capture information at three levels: Article, treatment, and adherence measurement method. Article level data included variables such as title, author, year of publication, client sample characteristics, clinician characteristics, treatment sites, and treatment settings. Treatment level data included variables such as therapy type, specific therapy name, and quality control methods presented. Adherence measurement method level data included variables such as the name of the adherence

measurement method/instrument, description of measurement development, reference trails for measurement development, reporting of adherence measurement results, a variety of psychometric variables, and variables reflecting the details of procedures used to obtain, code, and score adherence data.

For all variables except those reflecting titles and authors, forced choice responses were constructed that reflected presence or absence of the variable (e.g., yes/no), and, where appropriate, multiple choice responses (e.g., a single article could describe multiple treatment models and programs and each was listed as a choice to be endorsed; each type of statistical reliability was listed, and so forth). Text fields were available to allow coders to enter information that did not conform to one of the pre-defined response choices. Response choices included an option to indicate the availability of information needed to code a response. For example, the variable “Clinician Discipline” first provided a response choice to signal sufficiency of information reported in the article – not specified, partially specified, and completely specified. Then, eight response choices reflected different disciplines.

**Coders and coder training**—Four individuals, including the project coordinator (who had a master’s degree) and three research assistants (with bachelor’s degrees in mental health fields and basic familiarity with research design) were trained to review articles using the coding manual. Coder training included didactic training sessions, group discussion, independent review and coding of articles, and practice coding sessions. Didactic training sessions included (a) general introduction of the project aims and methods; (b) thorough review of the coding manual, and (c) trial coding on nine articles pre-selected to reflect different types of fidelity measurement methods. The codes supplied by coders on these nine trial articles were compared to codes independently assigned by the project investigators, and any discrepancies were discussed as a group. Questions that arose during training were resolved using a multi-step method. During training, coders recorded their questions (specifying article, page, variable, and question) and submitted these to the project coordinator (who also served as lead coder). The lead coder discussed the questions with an investigator. Both investigators conferred often to reach resolutions on these questions; resolutions were reported to the coding group and recorded for future reference. Questions that arose during coding were also resolved using a similar multi-step method.

Three coders and the study investigators were assigned an initial set of 20 articles to pilot test the reliability of the coding system. The ratings of each coder were reported on each item in the manual during consensus building meetings, and items with a lack of coder agreement were discussed. The coding manual was revised to specify with greater clarity the items prompting coder disagreement. Booster training sessions for coders were conducted using the revised manual. The 20 articles used in the pilot test of the coding process were then re-coded independently using the revised coding manual. Subsequently, a set of 40 articles was assigned to four coders (one new coder joined the project at this time) for the purposes of computing inter-rater reliability on the final version of the coding system. To assess inter-rater reliability of this system throughout the coding process, ten articles were randomly selected for coding by all four coders. Thus, inter-rater reliability of all coders on the final version of the coding system was computed on 50 (14.7%) of the 341 articles.

Inter-rater reliability amongst coders was calculated using Krippendorff’s Alpha (Hayes & Krippendorff, 2007). Krippendorff’s  $\alpha$  is rooted in data generated by any number of observers (rather than pairs of observers only), calculates disagreements as well as agreements, can be used with nominal, ordinal, interval, and ratio data, and with or without missing data points (Hayes & Krippendorff, 2007). Values of Krippendorff’s  $\alpha$  range from 0.000 for absence of reliability to 1.000 for perfect reliability. Reliabilities of .800 and higher (80% of units are perfectly reliable while 20% are the results of chance) are

considered necessary for data used in high stakes decision making (e.g., safety of a drug, legal decision, etc.). Variables with reliabilities between .667 and .800 are considered acceptable for rendering tentative conclusions about the nature of the variables coded (Krippendorff & Bock, 2007), and of .6 can be considered adequate for some scholarly explorations (Richmond, 2006). The current investigation is exploratory. We had no a priori hypotheses regarding the nature of adherence measurement methods for psychosocial treatments reported in peer-reviewed journals, and considerable uncertainty as to whether the measurement methods would be described with sufficient adequacy to allow for any type of coding at all. The values of Krippendorff's  $\alpha$  for variables included in the current analyses range from .615 to 1.0, with the mean of .788 for all coded items.

## Results

### Definition of Terms

Terms used to characterize psychosocial treatments are often used for disparate referents and purposes. Building on recent efforts to clarify such terms (Schoenwald, Garland, Southam-Gerow, Chorpita, & Chapman, 2011b), the following conventions are used in the present study. The term *model* refers to a broad, theoretically driven psychological treatment approach. As examples, cognitive behavioral therapy, interpersonal therapy, and parent training are models of therapy. The term *program* refers to a clearly defined psychological treatment from a particular model, generally specified in a manual. Parent Child Interaction Training (PCIT; Brinkmeyer & Eyberg, 2003) is an example of a specific treatment program. Specific instantiations of a program can be described as a protocol. A single program like PCIT may have multiple protocols associated with it (e.g., early vs. later versions of the manual, variations of the same manual as adapted for a specific population). Note that within the sample of articles reviewed for this study, the level of detail used to specify treatment models, programs, and protocols varies considerably.

Finally, we use the term “measurement method” to refer to the topic of primary interest in this review, namely how therapist adherence to a treatment was assessed. As reported in the following sections, sufficient information was available to ascertain the nature of some of the measurement methods and extent to which the scores generated by these methods were indeed reliable and valid for the reported use. For many others, however, the information presented described procedures for assessing adherence, but with insufficient detail and reference trails to discern the nature of the measurement method and availability of evidence for its reliable and valid use.

### Overview of Studies and Articles

This section provides an overview of the psychosocial treatment studies for which articles published between 1980 and 2008 reported efforts to assess therapist adherence.

**Articles and studies**—The 341 articles retained for this study reported on 304 distinct studies. Information from the vast majority (92.8%,  $n = 282$ ) of the studies was reported in a single article; information from 15 studies was reported in 2 articles, and information from a total of 7 studies was reported in 3, 4, or 6 articles.

**Treatment models implemented**—Cognitive behavioral treatments were implemented in over half (57.9%,  $n = 176$ ) of the studies. Motivational interviewing was deployed in 17.1% ( $n = 52$ ) of studies, and interpersonal therapy in 15.1% ( $n = 46$ ) of them. Family-based treatment models were implemented in 14.5% ( $n = 44$ ) of studies, and parent training models 10.2% ( $n = 31$ ) of studies. Psychodynamic and psychoanalytic treatment models were featured in 9.2% ( $n = 28$ ) of studies, and behavioral treatment models in 2% ( $n = 6$ ) of them.

A variety of treatments with unique descriptors that did not easily lend themselves to categorization were implemented in a total of 12.8% ( $n = 39$ ) of studies. Among these were treatments implemented in control conditions in randomized trials testing specific treatment models and programs, for which specific names were not always provided.

**Number of studies using the same adherence measurement method**—Across the 304 studies, 89% ( $n = 272$ ) reported on use of an identifiable adherence measurement method. Over half (58.5%,  $n = 159$ ) of studies reported use of a single, unique measurement method. Across the remaining 41.5% ( $n = 113$ ) of studies, the same adherence measurement method was used in more than one study, with frequency of method use ranging from 2 to 12 studies. A total of 249 measurement methods were identified. Of these, 23.7% ( $n = 59$ ) were identified with a title, and 76.3% ( $n = 190$ ) were not.

### Clinical Context of Adherence Measurement

This section focuses on the clinical context in which the 249 adherence measurement methods were used. Given our interest in illuminating the extent to which these methods have been used under conditions that resemble routine practice, the clinical context characteristics of primary interest were types of treatments, clinical problems, client populations, treatment settings, and clinicians providing treatment. Note that several types of treatment, and therefore several distinct adherence measurement methods, could be evaluated in the same study. Thus, the representation of adherence measurement methods reported in this section often exceeds 100% per category.

Of the 249 distinct adherence measurement methods used in the studies, eight measurement methods accounted for 21 unique entries (i.e., 21 of the 249 unique measurement methods actually represented multiple versions of eight measures). These measurement methods had been revised over time and/or for different clinical populations. Some revisions of items in the original instruments were made in successive studies with the same types of clinical problems and client populations. The titles of these measurement methods included phrases such as “revised,” or “version 1, version 2, version 3” and so forth. Among these measurement methods are the Therapist Behavior Rating Scale; (TBRS; Hogue, Liddle, Rowe, Turner, Dakof, & LaPann 1998); Yale Adherence and Competence Scales (YACS; Carroll, Connors, Cooney et al., 1998; Carroll, Nich, Sifry, Frankforter, Nuro, Ball, et al., 2000); Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon, Evans, Auerbach, DeRubeis, Elkin, Lowery, et al., 1988). Some instruments were revised when a treatment program was adapted for use with a new clinical problem or population; that is, when an additional protocol was specified for the program. An example is the instrument developed to index adherence to an adaptation of Multisystemic Therapy (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009) developed for substance abusing/dependent adolescents that incorporates Contingency Management. The adherence instrument for standard MST is the Therapist Adherence Measure – Revised (TAM-R; Henggeler, Borduin, Huey, Schoenwald, & Chapman; 2006); the instrument for the adaptation is MST TAM-CM (Chapman, Sheidow, Henggeler, Halliday-Boykins, & Cunningham, 2008). For current purposes, the sample size of adherence measurement methods remains 249 given that different versions of measures could have unique characteristics.

Table 1 summarizes the distribution of adherence measurement methods across treatments and clinical context variables, including clinical problems, clients, providers, and treatment settings.

**Treatment models and programs**—As shown in Table 1, the majority (59%,  $n = 147$ ) of the measurement methods indexed adherence to a cognitive behavioral model of treatment. Among the specific cognitive behavioral treatment programs for which adherence measurement methods were reported were Contingency Management (1.6%,  $n = 4$ ), Coping with Depression – Adolescents (1.2%,  $n = 3$ ), and Coping Power (.8%,  $n = 2$ ).

About 15% (15.3%,  $n = 38$ ), of measurement methods indexed adherence to the motivational interviewing model, and 9.6% ( $n = 8$ ) indexed adherence specifically to Motivational Enhancement Therapy programs.

Adherence to family based therapy models was the focus of 14.5% ( $n = 36$ ) of the measurement methods, and adherence to five specific treatment programs was evaluated. These programs were: Multidimensional Family Therapy (2.9%,  $n = 7$ ); Multisystemic Therapy (1.6%,  $n = 4$ ); Brief Strategic Family Therapy (1.2%,  $n = 3$ ); Functional Family Therapy (.8%,  $n = 2$ ); and Structural Strategic Family Therapy (.4%,  $n = 1$ ). A sixth program was identified but not reliably coded.

Adherence to Interpersonal Therapy was assessed by 14.1% ( $n = 35$ ) of the measurement methods, to psychodynamic and psychoanalytic treatments by 10% ( $n = 25$ ) of them, and to a parent training model by 13.3% ( $n = 33$ ) of them. Adherence was assessed in more than one study for the following specific treatment programs: Incredible Years (5.6%,  $n = 14$ ) Parent Child Interaction Training (3.2%,  $n = 8$ ), and the Oregon Social Learning Model (.8%,  $n = 2$ ).

**Clinical problems and client populations**—Clinician adherence was most frequently assessed for treatments of substance abuse (28.5%,  $n = 71$ ), anxiety disorders other than PTSD (27.3%,  $n = 68$ ) and mood disorders (22.9%,  $n = 57$ ). Almost fifteen percent (14.9%,  $n = 37$ ) of adherence measurement methods focused on treatments for youth with disruptive behavior problems and/or delinquency. Fewer than 10% of measurement methods were used to assess adherence to treatments for eating disorders and PTSD. Fewer than 5% focused on treatments for psychoses, personality disorders, ADHD, and autism spectrum disorders.

The adherence measurement methods were used in studies of treatment for client samples of varying ages, genders, and racial and ethnic backgrounds. Just over half of the measurement methods (51%,  $n = 127$ ) were used in studies treating adults (individuals aged 18 and older) only. Nearly 40% (39.8%,  $n = 99$ ) were used in studies treating children only. Very few (2.4%) were used in studies in which both adults and children participated in treatment. Finally, 6.8% ( $n = 17$ ) of the methods were used in studies of clients for which age was not reported. Over three-quarters (77.9%,  $n = 194$ ) of the measurement methods were used in studies involving both male and female clients, 7.2% ( $n = 18$ ) were used in studies involving only male clients, and 3.6% ( $n = 9$ ) in studies involving only female clients. Over half (57%,  $n = 142$ ) of the measurement methods were used in studies that included Caucasian clients, 35.7% ( $n = 89$ ) were used in studies that included African American clients, and 33.7% ( $n = 84$ ) were used in studies that included Latino/Hispanic clients 16.5% ( $n = 41$ ). Just under 10% (9.6%,  $n = 24$ ) of the methods were used in studies that included Native American clients, 7.2% ( $n = 18$ ) were used in studies with clients identified as Multi-Ethnic, and 27.3% ( $n = 68$ ) were used in studies including clients whose race or ethnicity was identified as “Other.” Because the studies in our sample spanned more than 25 years, and conventions for reporting on race and ethnicity have changed during that time, information was not sufficiently uniformly available to reliably code the ethnicity and race of Asian/Pacific Islander participants; nor to code finer grained distinctions in race and ethnicity such as, for example, Hispanic non-white, Hispanic white.

**Treatment settings**—Over half of the measurement methods (55%,  $n = 137$ ) were used in clinic settings, although 25.7% ( $n = 64$ ) were used in academic clinics. Almost one quarter -- 22.5% ( $n = 56$ ) -- were used in community clinics, 19.3% ( $n = 48$ ) were used in clinics described in insufficient detail to be coded as academic or community clinics, and 1.2% ( $n = 3$ ) were used in Veteran's Administration (VA) clinics. Schools were the settings in which 6.8% ( $n = 17$ ) of adherence measurement methods were used. An additional 10% ( $n = 25$ ) of measures were used in community settings described without sufficient detail to discern the type of setting (clinic, hospital, residential treatment facility, community center, etc.).

**Treatment providers**—The clinicians whose adherence to treatment was indexed by the 249 measurement methods represented a variety of professions, disciplines and levels of education. Over half of the measures (52.6%,  $n = 131$ ) assessed doctoral level professionals, and another 21.7% ( $n = 54$ ) were doctoral students. Over one-third of the measures (37.8%,  $n = 94$ ) assessed clinicians with a master's degree, 15.7% ( $n = 39$ ) clinicians with bachelor's degrees only, and 5.2% ( $n = 13$ ) clinicians with high school degrees only. Over half (54.6%,  $n = 136$ ) of the measures assessed psychologists, just over a quarter (20.1%,  $n = 50$ ), assessed social workers, and 16.1% ( $n = 40$ ) assessed psychiatrists. Clinicians from other disciplines such as counseling, education, and marriage and family therapy were represented in small numbers (.8% – 7.6%). Clinicians in unspecified allied and paraprofessional disciplines could not be reliably distinguished.

### Characteristics of Adherence Measurement Methods

The frequency with which distinct measurement methods were used, and details regarding data collection, rating, and scoring of adherence are presented next, and summarized in Table 2.

**Frequency of measurement method use**—Just over half (55%,  $n = 137$ ) of the 249 measurement methods were used once, to assess adherence to a single treatment model or program. Just over one quarter (28%,  $n = 70$ ) were used twice, 6.8% ( $n = 17$ ) were used 3 times, 4% ( $n = 10$ ) were used four times, three were used five times, and a total of 12 measures were used more than six times. The multiple uses of instruments occurred within the same studies, and/or across different studies over time. Just over one quarter (28.9%,  $n = 72$ ) of the instruments were used to index differentiation between treatment conditions in the same study. Some of these, and others, were used to assess adherence to the same treatment model or program across distinct studies.

**Methods used to obtain adherence data**—Data for nearly three-quarters (71.5%,  $n = 178$ ) of the instruments were obtained via *observational* methods. Of these 178 instruments, 56.2% ( $n = 100$ ) used audio recordings, 41% ( $n = 73$ ) video recordings, and 2.8% ( $n = 5$ ) required live observation of treatment sessions. For the majority (78.7%,  $n = 140$ ) of these observational measurement methods, select treatment sessions were coded for adherence (as opposed to all sessions). The proportion of sessions coded was not reported for over half (52.8%,  $n = 94$ ) of the instruments. For the remaining instruments, the proportion of sessions coded ranged from 7 – 62%, with over one quarter of the instruments (27%,  $n = 23$ ) using 20 – 25% of treatment sessions to code adherence. Most instruments (80%,  $n = 144$ ) required coding of entire treatment sessions, while a few (6.2%,  $n = 11$ ) required coding only segments of treatment sessions.

The individuals who provided ratings on the observational adherence measurement (i.e., “coders”) included clinicians (22.9%,  $n = 57$ ), study authors or experts in a treatment (20.5%,  $n = 51$ ) and university students or research assistants who were neither clinicians nor study investigators (26.9%,  $n = 67$ ). Information about coder training was provided for



15.2% ( $n = 27$ ) of the instruments. Reports of periodic refresher training for coders were provided for 9% ( $n = 16$ ) of instruments. Information about the amount of time required to code observational data and to calculate adherence scores after the data were coded was available for only four measurement methods. Scoring required 60 minutes for three of them, and 90 minutes for one of them.

Just over one-quarter (26.1%,  $n = 65$ ) of the 249 measurement methods were written forms completed by a variety of respondents, methods that can be categorized as *indirect* (Schoenwald et al., 2011a) because the means by which data are obtained are something other than direct observation of a session. Among these 65 instruments, the majority of respondents rating adherence was therapists (86.2%,  $n = 56$ ). Clients assessed therapist adherence for 16.9% ( $n = 11$ ) of the instruments, and clinical supervisors assessed adherence for one of them. Fewer than ten instruments were verbally administered to respondents or completed on the basis of case record reviews. All treatment sessions were rated for 66.2% ( $n = 43$ ) of the written instruments, selected treatment sessions for 12.3% ( $n = 8$ ) of them. Rating of sessions once weekly and once monthly occurred for fewer than 10% of the methods.

Information about how adherence data were scored –by hand versus computer scored–was provided for two instruments, one observational and scored by hand, the other written and scored using a computer. The amount of time required to score adherence data was reported for four observational instruments but could not be reliably scored. Adequate information about data collection and scoring to allow for the estimation of the costs of these activities was presented for only three (1.2%) measurement methods. Nine instruments (3.6%) were described as used routinely in practice.

**How was adherence rated?**—Based on the information provided in the published articles, it was difficult for the research team to discern the actual indicator of adherence used in the measurement methods. Types of rating pre-defined in the coding manual included occurrence (yes/no), frequency counts (counting how often a specific interaction, topic, behavior, occurs, expressed numerically), ratings (numeric rating regarding the extent to which something occurred that requires making a judgment), and size of the rating scale. Although some information about the nature of the ratings was reported in 62% ( $n = 154$ ) of the measurement methods, coder reliability of the information presented was low ( $\alpha = .506$ ). Accordingly, results are not presented here of our evaluation of the extent to which occurrence, frequency counts, and ratings were used to index adherence.

**Psychometric evaluation**—For nearly three-quarters (74.3%,  $n = 185$ ) of the measurement methods, the results of adherence measurement were reported in the articles. That is, articles presented data regarding adherence scores, and/or the proportion of sessions, clinicians, or both for which adherence score targets were met. Information about psychometric properties was reported for just over one-third (35%,  $n = 87$ ) of the 249 instruments. The psychometric information reported included statements regarding types and adequacy of validity and reliability evidence, nature of statistics used to evaluate various types of reliability and validity, and statements asserting the adequacy of these statistics.

The most prevalent means of assessing reliability of the scores obtained using the instruments was Intraclass Correlations (ICCs; 17.3%,  $n = 43$ ), Cohen's Kappa used with 5.2% ( $n = 13$ ), and Kendall's Coefficient of Concordance with 1.2% ( $n = 3$ ) attesting to the observational methods used to assess adherence. Cronbach's alpha (7.6%,  $n = 19$ ), and reliability assessment was not sufficiently described to allow for classification for 11.6% ( $n = 29$ ) of instruments.

Information about any validity evidence – concurrent, convergent, discriminant, predictive -- was presented for less than 5% of the adherence measurement methods. Associations between adherence scores and client outcomes were reported for only 10.4% ( $n = 26$ ) of the 249 instruments.

### Characteristics of Adherence Measurement Methods by Clinical Context Factors

To explore the relative use of adherence measurement methods across different types of clinical problems, populations, and treatments, we cross-tabulated select adherence measurement variables with treatment model, clinical problem area, and clinical population (i.e., child, adult, or both). For example, as reported previously, 34.9% ( $n = 87$ ) of all measurement methods had reported psychometrics. When examined by treatment model type, the distribution was 24.3% to 52%, with the rate of adherence measurement of almost all treatment model types falling within  $\pm 10\%$  points of the 34.9% mean. The two treatment model types outside this range were motivational interviewing (50%) and psychodynamic/psychoanalytic (52%). The rate of use of measurement methods with psychometrics reported across clinical problem areas ranged from 32.4% to 72.7%, with psychoses (72.7%) and substance abuse (46.5%) the two clinical problems outside  $\pm 10\%$  of the mean. The rate of use of measurement methods with psychometrics reported by age of clinical population ranged from 16.7% for measures used with both children and adults, to 29.3% for measures used with children only, and 37% for measures used with adults only. Thus no age group was outside the  $\pm 10\%$  range of the mean.

There was also some variability by clinical context on rates of use of adherence measurement methods in community settings. Overall, 32.5% ( $n = 81$ ) of the methods were used in a community setting (clinic, hospital, school, etc.). The rate of use of these methods varies as a function of treatment model, and ranged by model from 17.1% to 42.1%. Except for methods assessing adherence to interpersonal therapy (17.1%), community-based use of adherence measurement methods across all treatment models was within  $\pm 10\%$  points of the aggregate mean. The use of adherence measurement methods across clinical problem areas ranged from 22.2% – 59.5%, with methods to assess adherence to treatments for eating disorders more than 10% below the mean, at 22.2%. Methods indexing adherence to treatments for disruptive behavior problems/delinquency and psychoses were used in the community at higher rates, 59.5% and 54.5%, respectively. There was also a difference in rate of use of methods in community settings by clinical population, with methods assessing treatments for children demonstrating a higher rate of use in community settings (46.5%) compared to methods assessing treatment for adults (21.3%).

Finally, we assessed for differences in use of methods characterized by both scores for which psychometric properties were reported and use in a community setting. Only 15.2% ( $n = 38$ ) of measurement methods met both criteria. The range across treatment model types was 6% to 28%, with motivational interviewing (26.3%) and psychodynamic/psychoanalytic (28%) more than 10% points from aggregate mean. The range by clinical problem was 7.7% to 45.5% with psychoses (45.5%), and substance abuse (32.4%) as outliers with rates greater than 10% points of the mean. Finally, rates by clinical population (child, adult or both) ranged from 14.1% to 16.7%, thus no group fell outside the  $\pm 10$ -percentage point mean range.

## Discussion

The primary objectives of this study were to identify the measurement methods used to assess therapist adherence to evidence-based psychosocial treatments and the extent to which these methods are effective (i.e., have evidence for valid and reliable use of scores) and efficient (feasibly used in routine care). Notably, 249 distinct therapist adherence

measurement methods were identified. This number suggests considerable investigative attention has been directed toward the measurement of clinician adherence. The quality and yield of these measurement efforts with respect to effectiveness and efficiency was more difficult to discern on the basis of published information than we had anticipated. With respect to evidence of effectiveness, the results of our review were not terribly encouraging. Although numeric indicators of adherence (i.e., adherence scores, proportion of clinicians adherent) were reported for three-quarters of the measurement methods, psychometric properties of the scores were reported for just over one third of them, and evidence of predictive validity (evidence of meaningful associations between adherence scores and client outcomes) was reported for only ten percent of them.

### Clinical context of measurement use

With respect to the efficiency and feasibility of fidelity measurement methods in routine care, two types of information gleaned from our review are particularly pertinent to our consideration. (1) The extent to which the clients, clinicians, and clinical settings represented in the reviewed studies resemble those found in routine care settings; and (2) the extent to which the time, training, expertise, equipment, and materials needed to obtain, score, and report on adherence can be made available within the administrative, supervisory, and documentation practices of an organization (Schoenwald et al., 2011a). With respect to the former, the clinical populations and contexts in which in the adherence measurement methods were used resemble aspects of the populations and contexts in routine care, thus supporting the potential contextual fit and feasible use of the methods in routine care. For example, the most frequently targeted clinical problems assessed by the measurement methods are the highest prevalence problems in the community (Costello, Copeland, Angold, 2011; Ford, Goodman, & Meltzer, 2003; Kessler & Wang, 2008), specifically substance use (28.5%), anxiety disorders excluding PTSD (27.3%), mood disorders (22.9%), and disruptive behavior/delinquency (14.9%). The measurement methods were also used with clinical samples represented by a relatively balanced distribution by gender, race/ethnicity, and developmental age (child vs. adult), thus reflecting some of the diversity found in routine care. In addition, almost forty percent ( $n = 99$ ) of the methods were used with clinicians with a master's level education or less; and, over a quarter of the methods were used with clinicians from the disciplines highly represented in community care (e.g., social work, counseling, education, marriage/family therapy) (Peterson et al., 2001; Schoenwald et al., 2008).

The distribution of measurement methods across treatment models was, however, somewhat unbalanced, disproportionately assessing treatments not yet commonly implemented as intended in routine care. Specifically, fifty-nine percent of the measurement methods assessed the CBT model. Other evidence-based models, including motivational interviewing, interpersonal therapy, family therapy, and parent training were each assessed by fewer than 15% of the methods and were thus under-represented compared to CBT. Although limited, research on community practice with child and adult clinical populations suggests routine care clinicians often incorporate therapeutic techniques from multiple treatment models into their practice (Cook, et al, 2010; Garland et al., 2010). To assess which techniques clinicians incorporate and to what effects on client outcomes, validated methods of assessing adherence to these techniques are needed. The availability of already validated measurement methods for different treatment models and programs that index such techniques and thus could be used to evaluate the nature and impact of these routine practice patterns is uneven.

Almost one-third (32.5%) of the measurement methods were reportedly used in community clinical settings (clinics, schools or hospitals) and there were some notable differences by treatment model, clinical problem and client population in rates of use in these community

settings. Methods assessing adherence to interpersonal psychotherapy (for any clinical problem) and treatment for eating disorders had notably low rates of use in community settings. Alternatively, methods assessing adherence to treatments for disruptive behavior / delinquency and psychoses had notably high rates of community use. In addition, methods used with child clinical samples had a higher rate of use in the community (46.5%), compared to methods used with adults (21.7%). These findings highlight treatments for clinical problems and populations that require greater attention to adherence measurement methods, and those for which there is more promising evidence of availability and feasibility in community settings of current adherence measurement methods (e.g., methods assessing treatments for disruptive behavior or delinquency in children; Schoenwald et al., 2011b).

### Data collection and scoring

Almost one quarter of the observational measurement methods (19.1%) were used in community settings. Observational coding has long been considered the gold standard of psychotherapy adherence measurement, and has typically been a time and labor-intensive endeavor financed by research dollars (Schoenwald et al., 2011a). Unfortunately, very little of the information needed to estimate the time and cost of data collection and scoring was provided in the body of work reviewed for this study. Accordingly, we have no empirical basis for consideration of the extent to which the costs of different types of adherence measurement methodologies (i.e., observational, verbal, written), or of specific adherence measurement instruments, could be borne in routine care. Some impressions in this regard can be generated on the basis of the limited information available. For example, it appears 40 to 60 hours may have been required to train coders of 12 observational methods; however, the descriptions even of this seemingly straightforward construct were sufficiently variable that coder reliability was low ( $\alpha=.597$ ). On the one hand, the funds to support 40 hours (a standard work week in the U.S.) of coder training or more may not be available in current service budgets. On the other hand, if this investment yields valid and reliable coding of the behavior of multiple clinicians treating many clients, and the scores can be used to improve treatment implementation and outcomes, then the value proposition for clients, provider organizations, and payers may prompt inclusion of these costs in service reimbursement rates.

### Implications and future directions for research

The results of our review suggest there is considerable room for improvement with respect to the use of adherence measurement methods in routine care. Just as published evidence of the gap between psychotherapies as deployed in efficacy studies and community practice has catalyzed effectiveness and implementation research, so too our results suggest there is a gap that warrants bridging between adherence measurement methods devised for use primarily as independent variable checks in efficacy studies and those that can be used in diverse practice contexts. Indeed, a research funding opportunity announcement recently issued by the National Institute of Mental Health identifies the need to develop and test methods to measure and improve the fidelity and effectiveness of empirically supported behavioral treatments (ESBTs) implemented by therapists in community practice settings (National Institute of Mental Health, 2011). In the face of evidence of clinician effects on treatment integrity and outcomes, the announcement notes, “surprisingly little is known about how to extend effective methods of ESBT training and fidelity maintenance used in controlled studies to community practice” (NIMH, 2011, p.3), and that reliable, valid, and feasible methods of fidelity measurement are among the innovations needed to build the pertinent knowledge base.

The prospect of large-scale dissemination and implementation of effective psychosocial treatment holds great promise for consumers, clinicians, and third party purchasers. Absent

psychometrically adequate and practically feasible methods to monitor and support integrity in practice, this prospect also holds the potential to “poison the waters” for empirically supported treatment models, programs, and treatment-specific elements or components with demonstrated effectiveness. That is, without valid and feasible ways to demonstrate what was implemented with clients, the outcomes – good or bad – may be attributed to treatment that was not, in fact delivered.

## Limitations

Limitations in the conclusions that can be drawn from this study relate primarily to the reliance on data reported in the abstracted articles. Specifically, it is important to acknowledge that we can only report on the characteristics of studies and measurement methods that were reported in the articles. For example, if psychometric analyses were not reported in any of the articles in which a particular measurement method was used, the method would be coded as “no psychometrics reported.” Scant operational details of adherence measurement – how the data were obtained, scored, reported – were reported, so we know little about the resources (e.g., personnel, time, equipment, etc.) required to obtain, score, and report adherence. We do not know the extent to which such information exists but is not published, or instead remains to be collected anew. Similarly, scant details of the clinical context were reported, so we know little about the organizational (mission, policies and procedures, social context variables), case mix (types and numbers of clients treated), and fiscal characteristics likely to affect feasibility of using specific adherence measurement methods in these contexts.

The results of this review should also be interpreted in the context of the article sampling parameters, including for example, the time frame (1980–2008), and the exclusion of articles reporting on the treatment of health-related problems. With respect to the time frame, at least two phenomena may be pertinent to the nature and amount of the information authors reported about adherence measurement. (1) University-based efficacy studies were likely prevalent in the earlier decades of publication, with effectiveness and implementation studies only appearing more recently, such that measurement methods designed for the purposes of efficacy research dominated the sample. (2) Editorial guidelines have changed over time regarding manuscript length and contents, including the inclusion of pertinent information regarding adherence measurement methods. With respect to the latter, professional journal requirements to report on treatment fidelity have appeared relatively recently. For example, the *Journal of Consulting and Clinical Psychology* now requires authors of manuscripts reporting the results of randomized controlled trials to describe the procedures used to assess for treatment fidelity, including both therapist adherence and competence and, where possible, to report results pertaining to the relationship between fidelity and outcome. Accordingly, a review such as this conducted five years hence may reveal considerably more information regarding adherence measurement methods, and, potentially, greater use of them in community practice settings. Finally, although the average inter-rater reliability across all reported items was strong, a few items had only adequate reliability. Resource limitations prohibited evaluation of the extent to which the lower reliability of these items was a function of the limited information presented in the articles about them, the nature of the variable definitions used for coding the items, or alternative explanations. Evaluation of possible coder effects, however, found no evidence of individual coder bias.

## Conclusion

As we have suggested elsewhere (Schoenwald et al., 2011a), the purposes for which adherence measurement methods are used in routine clinical practice may differ from those for which the methods were originally developed. Measurement methods for high stakes

purposes such as decision making by third party purchasers regarding program funding or de-funding must be adequate to the task. The criteria for the adequacy of measurement methods – that is, requirements for consistency (reliability) and accuracy (validity) of the data derived from them -- may or may not be less stringent for some purposes than for others, for example, for quality improvement but not evaluative purposes. As fidelity measurement – including the measurement of therapist adherence -- moves from the realm of efficacy research to effectiveness and implementation research, and ultimately into implementation monitoring systems used in practice (Aarons et al., 2011), it will be important to develop evidence for the feasible, reliable, and valid use of adherence instruments for particular purposes – that is, to make specific types of decisions about therapist adherence – in practice contexts.

## Acknowledgments

The primary support for this manuscript was provided by NIMH research grant P30 MH074778 (J. Landsverk, PI). The authors are grateful to V. Robin Weersing for consultation regarding coding procedures for meta-analytic reviews and adherence to treatments for depression and anxiety, to Sharon Foster for consultation on diverse methods for evaluating the inter-rater reliability of multiple coders, and to Jason Chapman for his close attention to the rigor and language of measurement. The authors thank Emily Fisher for project management and manuscript preparation assistance, and William Ganger for the construction, management and evaluation of data files populated with hundreds of variables at multiple levels.

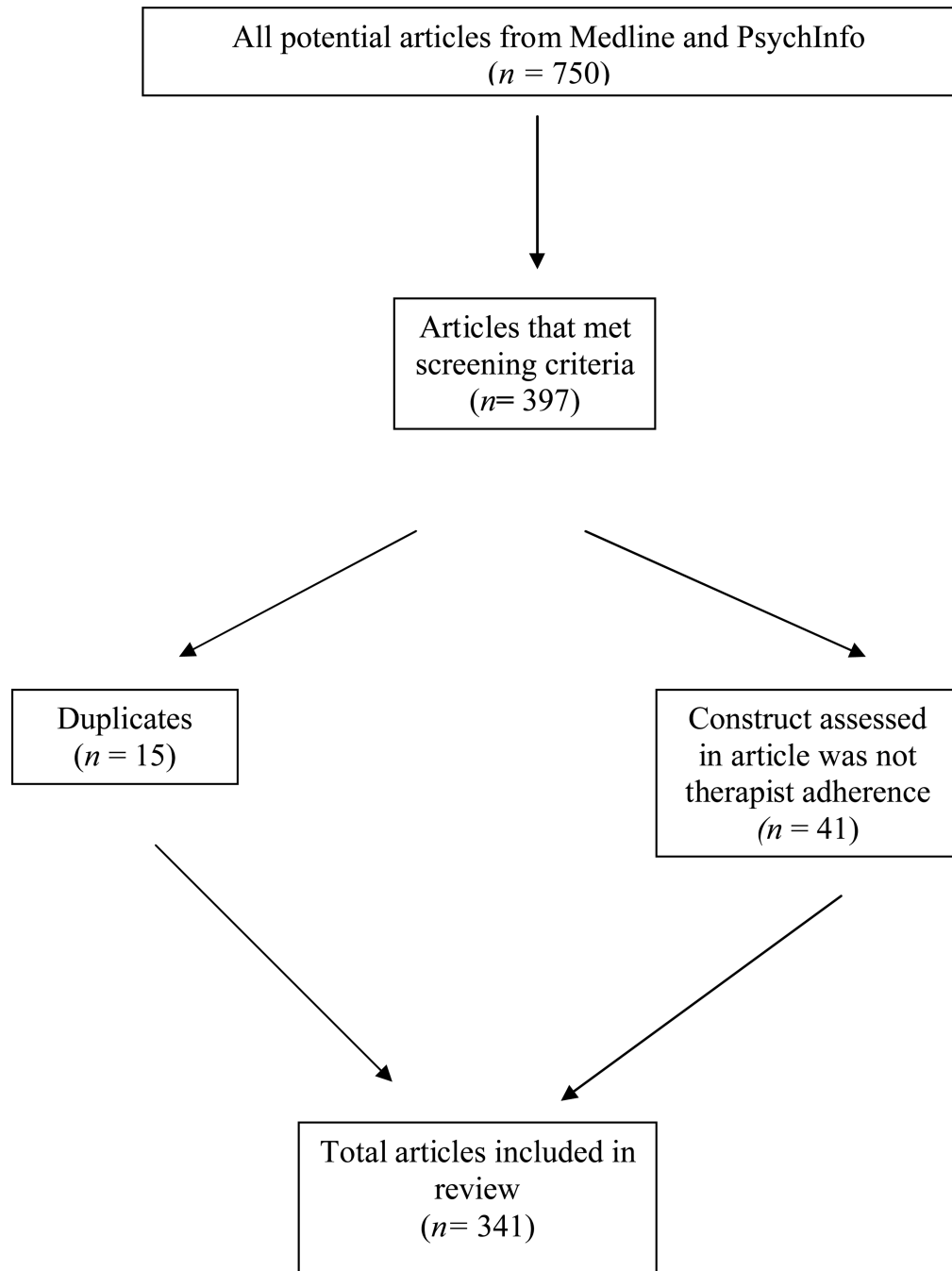
## References

- Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38:4–23. [PubMed: 21197565]
- Bergin, AE.; Garfield, SL., editors. *Handbook of Psychotherapy and Behavior Change*. Fourth Edition. New York: John Wiley and Sons, Inc; 1994.
- Brinkmeyer, M.; Eyberg, SM. Parent-child interaction therapy for oppositional children. In: Kazdin, AE.; Weisz, JR., editors. *Evidence-based psychotherapies for children and adolescents*. New York: Guilford; 2003. p. 204-223.
- Carroll KM, Connors GJ, Cooney NL, et al. Internal validity of project MATCH treatments: discriminability and integrity. *Journal of Consulting and Clinical Psychology*. 1998; 66:90–303.
- Carroll KM, Nich C, Sifry R, Frankforter T, Nuro KF, Ball SA, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*. 2000; 57:225–238. [PubMed: 10661673]
- Chambless DL, Ollendick TH. Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*. 2001; 52:685–716.
- Chapman JE, Sheidow AJ, Henggeler SW, Halliday-Boykins CA, Cunningham PB. Developing a measure of therapist adherence to contingency management: An application of the Many-Facet Rasch Model. *Journal of Child and Adolescent Substance Abuse*. 2008; 17(3):47–68. [PubMed: 20871744]
- Cook JM, Biyanova T, Elhai J, Schnurr PP, Coyne JC. What do psychotherapists really do in practice? An internet study of over 2000 practitioners. *Psychotherapy: Theory, Research, Practice, Training*. 2010; Vol. 47(2):260–267.
- Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? *The Journal of Child Psychology and Psychiatry*. 2011; 52:1015–1025. Epub 2011 Aug 4.
- Ford T, Goodman R, Meltzer H. The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2003; 42:1203–1211. [PubMed: 14560170]
- Garland AF, Brookman-Frazee L, Hurlburt MS, Accurso EC, Zoffness R, Haine RA, Ganger W. Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services*. 2010; 61:788–795. [PubMed: 20675837]

- Hayes SC. Market-driven treatment development. *The Behavior Therapist*. 1998; 21:32–33.
- Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*. 2007; 1:77–89.
- Hibbs, ED.; Jensen, PS., editors. *Psychosocial Treatments for Child and Adolescent Disorders: Empirically-Based Strategies for Clinical Practice*. Washington, DC: American Psychological Association; 1996.
- Hogue A, Liddle HA, Rowe C. Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy*. 1996; 33:332–345.
- Hogue A, Liddle HA, Rowe C, Turner RM, Dakof GA, LaPann K. Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. *Journal of Counseling Psychology*. 1998; 45:104–114.
- Hollon, SD.; Evans, MD.; Auerbach, A.; DeRubeis, RJ.; Elkin, I.; Lowery, A.; Kriss, M.; Grove, W.; Thason, VB.; Piasecki, J. Development of a system for rating therapies for depression: Differentiating cognitive therapy, interpersonal therapy, and clinical management pharmacotherapy. Twin Cities Campus: University of Minnesota; 1988. Unpublished manuscript
- Implementation Methods Research Group (IMRG; 2009). *Fidelity Measure Matrix Coding Manual, Version 2*. Child and Adolescent Services Research Center, Rady Children’s Hospital San Diego, CA: Author;
- Kazdin, AE.; Weisz, JR., editors. *Evidence –Based Psychotherapies for Children and Adolescents*. New York: The Guilford Press; 2003.
- Kendall PC, Chambless DL. Empirically supported psychological therapies [Special section]. *Journal of Consulting and Clinical Psychology*. 1998; 66:3–167. [PubMed: 9489258]
- Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annual Review Public Health*. 2008; 29:115–129.
- Krippendorff, K. Testing the reliability of content analysis data: What is involved and why. In: Krippendorff, K.; Bock, MA., editors. *The Content Analysis Reader*. Thousand Oaks: Sage Publications, Inc; 2007. p. 350-357.
- Liddle, HA.; Santisteban, DA.; Levant, RF.; Bray, JH., editors. *Family Psychology: Science-Based Interventions*. Washington, DC: American Psychological Association; 2002.
- Manderscheid RW. From many to one: Addressing the crisis of quality in managed behavioral health care at the millennium. *The Journal of Behavioral Health Services & Research*. 1998; 25:233–238. [PubMed: 9595886]
- Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*. *PLoS Med*. 2009; 6(6):e1000097. [PubMed: 19621072]
- National Institute of Mental Health. Funding Opportunity Announcement: Optimizing Fidelity of Empirically-Supported Behavioral Treatments for Mental Disorders (R21/R33). 2011. Retrieved from <http://grants.nih.gov/grants/guide/rfa-files/RFA-MH-12-050.html>
- Perepletchikova F, Treat TA, Kazdin AE. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*. 2007; 75:829–841. [PubMed: 18085901]
- Peterson, BD.; West, J.; Tanielian, TL.; Pincus, HA.; Kohut, J., et al. Mental health practitioners and trainees. In: Manderscheid, RW.; Henderson, MJ., editors. *Mental Health, United States*. Vol. 2000. DHHS Publication No. SMA 01-3537; 2001.
- Richmond, L. Intercoder Reliability. Paper presented at the First Annual Verbatim Management Conference; Cincinnati, OH: 2006 Sep.
- Schoenwald SK, Chapman JE, Kelleher K, Hoagwood KE, Landsverk J, Stevens J, Glisson C, Rolls-Reutz J. The Research Network on Youth Mental Health. A survey of the infrastructure for children’s mental health services: Implications for the implementation of empirically supported treatments (ESTs). *Administration and Policy in Mental Health and Mental Health Services Research*. 2008; 35:84–97. [PubMed: 18000750]
- Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011a; 38:32–43. [PubMed: 20957425]

- Schoenwald SK, Garland AF, Southam-Gerow MA, Chorpita BF, Chapman JE. Adherence measurement in treatments for disruptive behavior disorders: Pursuing clear vision through varied lenses. *Clinical Psychology: Science and Practice*. 2011b; 18:331–341. [PubMed: 22563149]
- Silverman WK, Hinshaw SP. Evidence-based psychosocial treatments for children and adolescents: A ten year update. [Special issue.]. *Journal of Clinical Child and Adolescent Psychology*. 2008; 37(1)
- Special Issue on Empirically Supported Psychosocial Interventions for Children. *Journal of Clinical Child Psychology*. 1998; 27(2)
- Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology*. 2009; 77:595–606. [PubMed: 19634954]
- Weisz, JR.; Kazdin, AE., editors. *Evidence –Based Psychotherapies for Children and Adolescents, Second Edition*. New York: The Guilford Press; 2010.
- Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting & Clinical Psychology*. 1981; 49:156–167. [PubMed: 7217482]





**Figure 1.**  
Flow of Selection and Retention for Review of Published Articles

**Table 1**

Clinical contexts in which adherence measurement methods were used (n = 249)

Clinical Contextual Variables	n of measurement methods	% of methods used with contextual variable <sup>1</sup>
<b>Treatment Models Assessed by Adherence Measurement Methods</b>		
Cognitive Behavioral	147	59.0%
Motivational Interviewing	38	15.3%
Family Therapy	36	14.5%
Interpersonal Psychotherapy	35	14.1%
Parent Training	33	13.3%
Psychodynamic/Psychoanalytic	25	10.0%
<b>Clinical Problems Assessed</b>		
Substance Use	71	28.5%
Anxiety (w/out PTSD)	68	27.3%
Mood Disorders	57	22.9%
Disruptive Behavior Problems & Delinquency	37	14.9%
Eating Disorders	18	7.2%
PTSD	13	5.2%
Psychoses	11	<5%
Personality	4	<5%
ADHD	4	<5%
Autism Spectrum Disorders	2	<5%
<b>Client Characteristics</b>		
<b>Client Age</b>		
Adults	127	51.0%
Children only	99	39.8%
Both Adults and Children	6	2.4%
<b>Client Gender</b>		
Males & Females	194	77.9%
Males only	18	7.2%
Females only	9	3.6%
<b>Client Race/Ethnicity</b>		
Caucasian	142	57%
African American	89	35.7%
Latino/Hispanic	41	16.5%
Native American	24	9.6%
Multi-Ethnic	18	7.2%

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/pubs/journals/pas](http://www.apa.org/pubs/journals/pas)

Clinical Contextual Variables	n of measurement methods	% of methods used with contextual variable <sup>1</sup>
Other	68	27.3%
<b>Treatment Settings</b>		
<b>Clinic Setting</b>	137	55%
Academic Clinics	64	25.7%
Community Clinics	56	22.5%
VA Clinics	3	1.2%
Could not be determined	48	19.3%
<b>Schools</b>	17	6.8%
<b>Community setting –unable to specify</b>	25	10%
<b>Treatment Provider Characteristics</b>		
<b>Provider Degree</b>		
PhD/MD	131	52.6%
PhD Students	54	21.7%
Master's degree	94	37.8%
Bachelor's degree	39	15.7%
High School diploma	13	5.2%
<b>Provider Discipline</b>		
Psychology	136	54.6%
Social Work	50	20.1%
Psychiatry	40	16.1%
Counseling	19	7.6%
Education	6	2.4%
Marital & Family Therapy	2	.8%

<sup>1</sup> Percentages within categories often total more than 100% since one adherence measurement method may be used in multiple clinical contexts within and across studies.

---

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/pubs/journals/pas](http://www.apa.org/pubs/journals/pas)

Table 2

## Measurement Method Characteristics

Measurement Method Characteristics	n of measurement methods	% of methods
<b>Type of Measurement Method (n =249)</b>		
Observational	178	71.5%
Written	65	26.1%
<b>Psychometric Evaluation</b>		
Results reported	185	74.3%
Psychometric properties reported	87	35%
<b>Reliability evidence provided</b>		
Intraclass Correlations	43	17.3
Kendall's Coefficient of Concordance	3	1.2%
Cronbach's alpha provided	19	7.6%
Cohen's Kappa	13	5.2%
<b>Validity evidence</b>		
Convergence, concurrent, or discriminant	6 – 11	2.4 – 4.4%
Association of adherence with outcomes	26	10.4%
<b>Details of Observational Data Collection (n = 178)</b>		
Audio recording	100	56.2%
Video recording	73	41%
Live observation	5	2.8%
<b>Treatment Session Sampling Strategy</b>		
Not reported	94	52.8%
Proportion of treatment sessions coded	140	78.7%
20 – 25% of sessions	23	27%
Entire session coded	144	80%
Session segments coded	11	6.2%
<b>Coding Personnel</b>		
Clinicians	57	22.9%
University students or research assistants	67	26.9%
Study authors or treatment experts	51	20.5%
<b>Details of Written Data Collection (n=65)</b>		
<b>Treatment Session Sampling Strategy</b>		
All sessions	43	66.2%
Once weekly	6	9.2%
Once monthly	4	6.2%
<b>Respondents</b>		
Therapists	56	86.2%
Clients	11	16.9%

<sup>1</sup> Percentages within categories often total more than 100% since one adherence measurement method may be used in multiple clinical contexts within and across studies.