



Keeping the Faith While Keeping It Real: Practical, Empirical Approaches to Evaluating Treatment Fidelity

Suzanne E. U. Kerns^{1,5} · Cameron M. Perrine² · Georganna Sedlar³ · Roselyn Peterson⁴ · Maria Monroe-DeVita³

Received: 1 December 2020 / Accepted: 26 April 2021 / Published online: 14 May 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Practical approaches towards fidelity measurement are critical for large-scale uptake of evidence-based practices (EBPs). This descriptive review summarizes approaches to evaluation of EBP fidelity that are empirically tested and implemented in “real-world” settings. The systematic literature search covered a 30-year period (1988–2018) and included published papers describing treatment fidelity strategies for psychosocial interventions. Characterization of articles included “scientific-empirical,” “practical-empirical,” or “practical non-empirical.” Articles characterized as “practical-empirical” were the focus of the review. Twenty-five articles met inclusion criteria for “practical empirical.” These articles were coded for fidelity data collection methods, the source of fidelity information, who arbitrates or determines fidelity, if the approach is at the practice or program level, if the approach is for a specific EBP or is generic, and if fidelity is measured at the individual practitioner versus team level. The type of empirical evidence provided and the extensiveness of fidelity domains assessed was characterized for each study. Results indicated a wide variety of fidelity measurement approaches. Most commonly, outside experts assessed treatment fidelity. Half of the approaches examined fidelity at the practice level and half at the broader program level. A similar ratio examined fidelity at the individual practitioner level compared with the team level. Most approaches were specific to particular interventions. The vast majority focus on the adherence subdomain of fidelity. To date, there is no single predominant approach to fidelity assessment in real-world settings.

Keywords Practical fidelity · Fidelity measurement · Treatment fidelity · Implementation · Literature review

For psychosocial interventions in mental health, delivery of treatment interventions as designed (i.e., fidelity) is a key implementation factor that ensures effective transportability

of interventions from highly controlled scientific studies into community-based settings (Fixsen et al., 2005; Proctor et al., 2011; Tabak et al., 2012). Fidelity is a critical implementation outcome (Proctor et al., 2011) that is directly associated with clinical outcomes (Cuddeback et al., 2013; McGuire et al., 2016; Schoenwald et al., 2003) and program cost effectiveness (Rollins et al., 2017). Because of this and the accountability that fidelity tracking provides, funders and policy makers increasingly require documentation of fidelity (Bellg et al., 2004; Schulte et al., 2009). However, most established methods for assessing and supporting treatment fidelity have origins in establishing internal validity within research studies (Bond & Drake, 2020) and not for widespread use in community-based settings. The feasible or practical translation of fidelity measurement in community-based settings is understudied (Suhrheinrich et al., 2019). The purpose of this review is to collate and assess the extent of practical and empirically supported approaches to fidelity measurement in community-based settings (Fig. 1).

✉ Suzanne E. U. Kerns
Suzanne.Kerns@cuanschutz.edu

¹ Center for Effective Interventions, Graduate School of Social Work, University of Denver, 2148 S. High Street, Craig Hall, Denver, CO 80208, USA

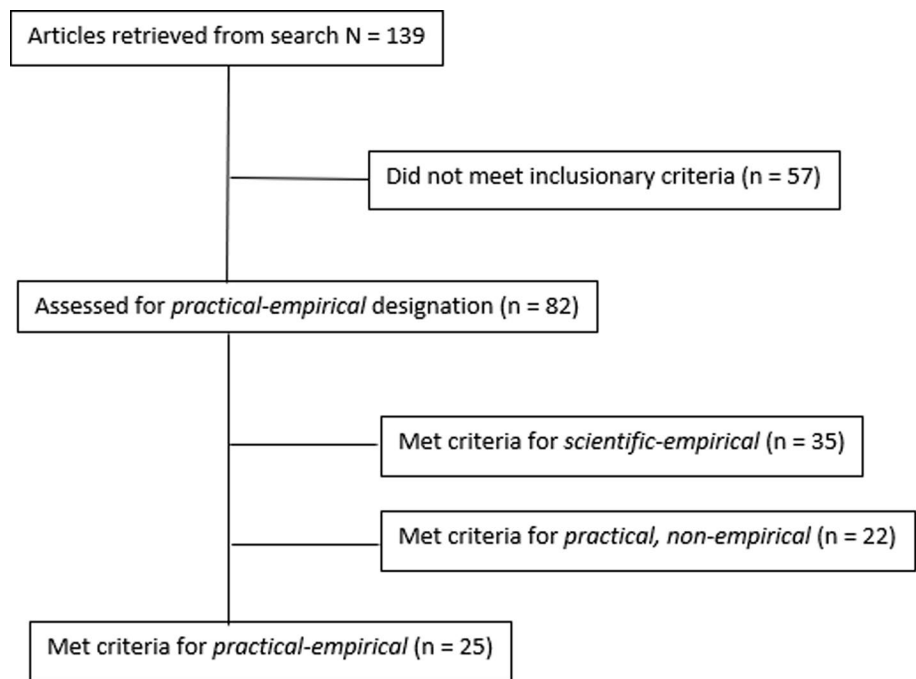
² Department of Psychological Science, University of Arkansas, 216 Memorial Hall, Fayetteville, AR 72701, USA

³ Department of Psychiatry and Behavioral Sciences, University of Washington, 1959 NE Pacific Street, Box 356560, Seattle, WA 98195, USA

⁴ Department of Psychology, University of Central Florida, 4111 Pictor Lane, Ste. 323, Orlando, FL 32816, USA

⁵ Kempe Center, University of Colorado School of Medicine, Aurora, USA

Fig. 1 PRISMA diagram



Fidelity tracking methods designed for research trials frequently rely on specific technological supports, require unique data entry, and have other features that are difficult to implement within community-based settings. In community-based contexts, such approaches may be impractical and construed as overly prescriptive, expensive, and labor intensive. A recent mixed-methods case study of behavioral health agencies highlights that, despite perceptions that fidelity measurement may benefit clinical practice, fidelity measurement can be viewed as extra and artificially constructed (e.g., use of audiotapes during sessions) within community-based settings and can be difficult to maintain over time (Barwick et al., 2020). Further, existing empirically supported practices often fail to provide fidelity supports and, when they do, these supports are rarely required (Rolls-Reutz et al., 2020). Without having practical and empirically supported methods for assessing fidelity in community-based settings, it is difficult to know whether implementation is on track and whether it will yield anticipated outcomes (Schoenwald et al., 2011).

Mental health agencies and practitioners indicate a desire for practical fidelity measurement tools (Kimber et al., 2019). In order to meet this need, some community-based treatment settings develop ad hoc strategies for establishing fidelity (often to meet funder requirements), yet these strategies are largely untested or not evaluated. Moreover, the most common strategy, use of a fidelity checklist by individual practitioners, has shown generally poor reliability compared to objective observational methods (Hurlburt et al., 2010). The purpose of this review is to describe and synthesize the available literature of fidelity measurement

of empirically supported practices (EBPs) in real-world settings and summarize the practical fidelity approaches that have some empirical basis. We focus this review on those fidelity support practices evaluated for psychosocial interventions for mental health and substance use disorders, and within the practice settings in which implementation is intended. Improving understanding of the status of practical fidelity supports furthers identification of which fidelity measurement practices are beneficial under which circumstances and highlights gaps in implementation science knowledge. Ultimately, moving towards practical and empirically supported fidelity measurement is a step toward improving implementation and sustainment of EBPs in community-based settings.

Definition of Treatment Fidelity

While there is no single definition of treatment fidelity, in summary, it is the delivery of treatment components as intended (McLeod et al., 2013). The literature on treatment fidelity measurement is inclusive of the individual practitioner level and the program or team level. The appropriate fidelity measurement tool is based on the unit of analysis. Practitioner-level fidelity tools focus specifically on clinician, therapist, or interventionist behavior within the context of delivering the program or practice. Program- or team-level fidelity measures can also examine individual practitioner behavior, despite typically at an aggregated level. Program-level fidelity may include macro-level fidelity elements such as program structure, target population, caseload size,

and staffing (Bond et al., 2000). Most program-level fidelity tools are stand-alone measures that focus on that larger unit of analysis, but some integrated fidelity tools assess both individual practitioner behavior and program features all in one scale (Teague et al., 2012). In absence of an integrated measure, some fidelity measurement approaches combine two different measures assessing each level of analysis (e.g., using the Illness Management and Recovery Fidelity Scale [IMR-FS; Gingerich & Mueser, 2010] to assess program structure and aggregated practitioner adherence to the IMR intervention as well as the IMR Treatment Integrity Scale [IT-IS; McGuire et al., 2012] to assess individual practitioner adherence to IMR practice).

At the practitioner level, the two most commonly assessed components of treatment fidelity are practitioner *adherence* and practitioner *competence* (Webb et al., 2010). Other less frequently assessed but important characteristics of fidelity are treatment *differentiation* (Schoenwald & Garland, 2013) and *relational factors* (Hukkelberg & Ogden, 2013). Briefly, *adherence* is how closely the delivered treatment matches the original model. Adherence compares what a practitioner does in a particular session with a client and examines over time whether the practitioner delivered all of the intervention components. *Competence* refers to the quality of treatment delivery. Theoretically, it is possible for a practitioner to “go through the motions” of a treatment manual but not deliver the treatment in a skilled, engaging, and motivating manner. Thus, assessing competence allows for a deeper examination of the quality of model delivery. *Differentiation* addresses the questions: to what extent is the delivery of the treatment distinct from other services (e.g., non-manualized components)? Did the practitioner add content that was not part of the intervention? Finally, *relational factors* refer to the alliance between a client and their practitioner, and the extent to which the client is involved in the treatment (McLeod et al., 2013).

Although each of these aspects of fidelity is widely considered critical components, there may be differential importance as it relates to client outcomes. A recent meta-analysis of child and adolescent psychotherapy studies revealed that the relationship between different aspects of fidelity and outcomes differs by fidelity domain (Collyer et al., 2019). For example, there was a small but significant relationship between adherence and outcomes. However, there were more mixed findings related to the relationship between competence and clinical outcomes (Collyer et al., 2019).

Assessment of Treatment Fidelity

Fidelity measurement occurs through direct and/or indirect methods. Direct methods include observation of practitioners delivering the treatment through live observation or

through review of audio or video recordings or conducting site visits to observe team functioning. Direct methods of evaluating fidelity typically include creating a coding manual to evaluate certain parts of the treatment process or document existence of required programmatic elements. Coding manuals may be inclusive of multiple aspects of fidelity, as described above. For example, the Therapy Process Observational Coding System-Revised Strategies (TPOCS-RS) (McLeod et al., 2015) assesses intervention integrity and differentiation. Indirect methods include self-report of what happened in session by the practitioner or client, use of a checklist completed post-session, or review of other supplemental treatment products such as case notes (Breitenstein, Fogg, et al., 2010; Breitenstein, Gross, et al., 2010). Indirect methods may include agency reporting on program characteristics or metrics at the team level. A third method is asking clients directly about their treatment experience. For example, MST includes a Therapist Adherence Measure-Revised (TAM-R) administered telephonically to clients once per month (Henggeler et al., 2006). Direct methods are the gold standard for measuring fidelity; however, these methods are not easily implemented in all settings, particularly community mental health agencies. Audiotaped sessions require HIPAA-compliant technological resources and practitioner time to upload sessions for review. Further, this approach is time intensive from the perspective of the fidelity reviewer who takes the time to evaluate and provide feedback to the practitioner and/or supervisor. Collection of data from clients is also labor intensive, such as the case with MST, and may necessitate coordination with a specialized call center and associated costs. If gold-standard approaches were utilized across all community-delivered EBPs, especially when practitioners are implementing more than one EBP at a time, fidelity assessment may become unwieldy, time intensive, and costly. However, indirect methods may be less reliable.

As fidelity measurement moves from the realm of research studies into practical applications in community-based settings, there have been attempts to make collection of fidelity data more efficient and less burdensome. However, assessing fidelity with more feasible and practical methods for the user comes with its own set of challenges. For example, the considerable variability in sources of routinely collected administrative data in community-based settings makes systematic collection across sites a major challenge. Instruments, such as practitioner-reported checklists, may not be used systematically or routinely (Schoenwald et al., 2011). Practitioner report does not always match with what actually happens in treatment (Hurlburt et al., 2010), and there may be important differences in the accuracy of self-reported fidelity across different types of treatment (Hogue et al., 2014). In addition, while practitioner-report checklists may be the most practical to implement, checklists typically

focus on adherence and rarely capture all four aspects of fidelity. For example, a systematic review of fidelity strategies for complex behavior change interventions found that less than half of the fidelity strategies included assessment of program delivery and the relational aspects of engagement (Walton et al., 2017).

Thus, there is a need to identify practical strategies for individual practitioners, supervisors, and administrators within provider organizations to measure fidelity that have some empirical basis. Tailored fidelity supports have the potential to help guide supervision practice. For example, a novel randomized control study explored the impact of symptom and fidelity monitoring and behavioral rehearsal on treatment fidelity to Trauma-Focused Cognitive-Behavioral Therapy (Dorsey et al., 2013). These strategies were selected because of their potential utility and feasibility in usual care settings. However, despite the potential relevance of such strategies, such approaches to fidelity monitoring rarely occur naturally within community-based mental health (Beidas et al., 2014; Dorsey et al., 2017). The extent to which fidelity assessment approaches, other than the traditionally gold-standard ones, have utility in real-world settings is an area of current debate.

Current Study

The current study aims to describe the state of the science with regard to practical, empirically supported fidelity assessment approaches. This review focuses on the status of fidelity assessment within community-based settings and identifies the practical, empirically supported fidelity tools within the literature (*practical-empirical fidelity measurement*). *Practical-empirical fidelity measurement* is operationalized as fidelity monitoring or assessment approaches and tools that have the potential for use across different implementation settings (i.e., not just a singular setting), are expected to be feasibly administered within real-world constraints (e.g., time, resource, etc.) and have some empirical evidence that they capture fidelity (either adherence or competence). Identifying *practical-empirical* approaches to evaluating and measuring treatment fidelity is challenging because few studies explicitly frame practical fidelity measurement and, as such, practical fidelity approaches are discussed within articles that focus on transportability or effectiveness research. Nevertheless, this review endeavors to synthesize articles that illustrate a range of practical applications of fidelity measurement. This review study is a first step towards articulating what is currently available and what gaps remain to make high-quality fidelity assessment practical and feasible to support implementation of empirically supported practices. The primary research questions are “What is the scope of practical approaches to

fidelity measurement in the extant literature?” and, “What are the primary qualities of practical, empirically supported approaches to fidelity measurement to date?”

Methods

Operational Definitions

The focus of this review is to describe practical, empirically examined fidelity measurement for psychosocial interventions and/or prevention programs delivered in community-based settings (i.e., community mental health centers, schools, home-based services; and not including hospitals, university-based settings, or specialty clinics). Fidelity assessment is conceptualized along two dimensions: *practical* versus *scientific* and *empirically tested* versus *not empirically tested*. No articles could be scientific and not empirically tested. Thus, articles were coded into one of three categories described below. Coded articles were deemed *practical* if the focus of the fidelity assessment was to evaluate treatment fidelity in real-world settings. This was distinguished from articles that provided fidelity assessment approaches strictly for the purposes of establishing internal validity within research studies (i.e., *scientific*). Articles were coded as *empirical* if there was any explicit examination of the qualities of the fidelity tool (e.g., reliability, validity, acceptability). *Non-empirical* articles were those that described a fidelity approach but did not examine its effectiveness or relationship between fidelity and outcomes. Thus, the three categories of studies emerging from this conceptualization are defined below.

Practical-Empirical

Study descriptions that (1) delivered the intervention and fidelity tools in “real-world” settings (i.e., not within the context of a highly controlled clinical trial) and (2) specifically evaluated any aspect of the fidelity measurement approach. This could include but was not limited to, the validity, acceptability, and/or clinical- or practitioner-level outcomes associated with the fidelity measurement approach. Tests of specific fidelity measures and measurement frameworks were included in this category as long as they were conducted in real-world settings. Fidelity evaluation approaches could be qualitative or quantitative in nature.

Articles coded as *practical-empirical* are the explicit focus of this review.

Practical Non-Empirical

Articles coded as *practical non-empirical* included (1) articles that focused on fidelity assessment guidelines,

frameworks, strategies, and tools in which use is described for community-based settings, or (2) studies of interventions in real-world settings that used fidelity measurement but did not evaluate the fidelity approach in any way. Descriptions of fidelity measurement approaches without a specific examination of any empirical aspects of the fidelity measurement tool or approach were included in this category. Often the approaches described had high face validity but were not scientifically evaluated. Alternatively, an effectiveness study carried out in a community-based mental health center that described use of a fidelity measurement approach but provided no information specific to how the fidelity approach performed within the context of the study was included in this category.

Scientific-Empirical

Studies that examined fidelity strictly within the context of a clinical trial without evaluating the practicality, feasibility, or acceptability of the fidelity approach in “real-world” settings (e.g., clinics, schools) were considered *scientific empirical*. Typically, these examinations of fidelity served the purpose of establishing internal validity for research studies. Within this category, the extent to which the strategies have utility in community-based settings may be implied but is not expressly addressed.

Search Strategy and Selection

The search was limited to the 30-year span from 1988 to 2018. There was only one article meeting our inclusion criteria that predated 1996. Thus, this period likely is comprehensive of the period when treatment fidelity emerged as a construct of interest outside of clinical trials. One hundred and thirty-nine articles were identified from PubMed, PsychINFO, and Google Scholar. Articles were initially selected based on the following keywords: Fidelity, Guideline Adherence, Treatment Adherence, EBPs, Implementation, Intervention (adults and children/youth), Programs and Practices, Guidelines, Quality Improvement, Quality Assurance, Treatment Integrity, Strategies, Health Care, Health Plan Implementation, Community Mental Health Services, Evidence-Based Medicine, Treatment Outcome, Program Evaluation, Quality Improvement Strategies, Patient Outcome, Progress monitoring, Competence/Competencies, and Therapist Competence. Publications from experts in the field were solicited, as well as citation/reverse citation searches of seminal articles. The inclusion criteria included articles explicitly examining fidelity approaches for evidence-based psychosocial interventions. Excluded publications included those that described the status of treatment fidelity within a larger implementation/research context (for example, Barber et al., 2007; Hoagwood, 2013; Kendall & Beidas, 2007).

Studies that examined the empirical support for treatment fidelity measures but did so outside of the intended use setting were not included. For example, Southam-Gerow et al. (2016) established the psychometric properties of the Cognitive-Behavioral Therapy Adherence Scale for Youth Anxiety by using doctoral student coders (Southam-Gerow et al., 2016). Also articles were excluded, which had fidelity as a secondary variable/outcome within a larger study (for example, Durlak & DuPre, 2008; Isett et al., 2007). Articles that described overall training of providers in EBPs were excluded if they did not have a specific focus on evaluating fidelity approaches (e.g., Beidas et al., 2015).

Coding Categories

All articles that met inclusionary criteria for *practical and empirical* were coded. Codes characterized the type of empirical data used for the fidelity approach, how fidelity data were collected, from whom, who determined if fidelity was met, at what level fidelity was measured, if the approach was designed for a specific EBP or not, if fidelity assessment happened at the individual or team level, and what fidelity domains were included (e.g., adherence, competence). An iterative process determined categories within each domain and new categories emerged as studies were evaluated to ensure the review comprehensively considered the range of approaches used. The study team established preliminary categories based on article review, research, and clinical experience. During the coding process, group discussion, and reconciliation resulted in revision or expansion of some categories. The final categories within each domain are described in more detail below.

Type of Empirical Data

Empirical data included the strategies used to determine the psychometrics or empirical support for the fidelity approach. These included reliability (e.g., measure, interrater), validity (e.g., measure, concurrent, predictive), pragmatism/feasibility, benchmarking, variability, normality, racial disparities, and comparative analysis. If applicable, the extent to which the fidelity approach demonstrated an association with clinical outcomes was indicated. Studies often included multiple strategies for establishing the empirical basis of the fidelity approach; thus, strategies are not mutually exclusive.

EBP Specific Versus EBP Agnostic/multiple EBPs

Fidelity approaches developed for a specific empirically supported practice, intervention, or treatment were coded as EBP specific. Approaches applied across a range of EBPs or that were not designed for a specific EBP were coded as EBP agnostic/multiple.

Data Collection Strategies

Data collection domains included observation (direct or video/audio), chart data (e.g., progress notes), billing data/service data (e.g., number of sessions, treatment dose information, etc.), practitioner-reported checklists/ratings, client outcomes, or interviews (in person, phone).

Data Collection Sources

The source of the fidelity information was coded as client report (clients or recipients of the intervention provide information related to program fidelity), practitioner report (practitioner or interventionists providing the intervention provide information related to program fidelity), supervisor report (clinical supervisors provide program fidelity information), and/or administrative report (program fidelity information is obtained through administrative records, which could include program notes, billing, other records).

Fidelity Arbitration/Determination

Fidelity arbitration is how and by whom a determination of program or practitioner fidelity is made. For each practical, empirical article, the individuals responsible for determining fidelity were categorized as one of the following: an outside expert (including an EBP purveyor or other individual not part of the EBP team), administrator within the organization, supervisor within the organization, or someone else.

Practice Level versus Program Level

Distinctions were made between those fidelity approaches that focused on the interactions between the practitioner and client (practice level) and program- or team-level factors that are necessary for implementing the intervention with fidelity, such as adhering to required caseload sizes and practice parameters (program level). These categories were not mutually exclusive. A fidelity approach could contain elements of both. For example, a fidelity assessment approach that consisted of therapists submitting audiotapes of sessions for coding by a treatment expert was coded as “practice level.” A fidelity assessment approach that consisted of a site visit in which information about the functioning of the program (e.g., size of the team, average caseloads, supervision structure, and assessment of agency policies and procedures as it relates to the program) was coded as “program level.”

Level of Fidelity Assessment (Practitioner vs Team)

Fidelity approaches designed to assess the fidelity of a specific practitioner (individual practitioner) were coded compared to those approaches that examined fidelity for an entire

clinical team (team level). These two categories were not mutually exclusive if the fidelity approach blended across levels. Specifically, when each practitioner delivering the EBP received their own fidelity assessment rating, the EBP was coded as “practitioner level.” If assessment ratings were provided collectively for the entire team or program, then the EBP was coded as “team level.” When individual level fidelity ratings were available and also rolled up into team-level scores, the EBP was coded as “both.”

Fidelity Domains Assessed

The extent to which the fidelity tool measured the following fidelity domains was coded: adherence (at the practitioner and/or organizational-level), competence, differentiation, and relational factors (McLeod et al., 2013). Note, neither differentiation nor relational factors were observed during coding.

The coding team (all authors) divided the articles and two coders provided ratings for every article. The entire coding team reviewed the article to resolve any discrepancies.

Results

Articles Meeting Inclusion Criteria

After culling the articles that did not match the inclusion criteria, eighty-two articles met criteria for *practical-empirical*, *practical non-empirical* or *scientific-empirical* approaches to evaluating treatment fidelity. Of those, thirty-five articles were determined to be *scientific-empirical* articles and 22 were *practical, non-empirical* articles. This left 25 *practical-empirical* articles for inclusion in this review. These articles were all published from 1998 to 2018 and constitute the focus of the current review.

Practical Empirical Approaches

Please see the online supplementary table for an overview of the articles that were determined to be *practical-empirical*, and are the subject of this review.

Establishing the Empirical Basis for the Fidelity Approach

Substantial variability was observed with regard to the empirical strategies used within identified studies (see Table 1). It was most common that some measure of reliability and validity were used to assess the fidelity tool ($n = 16$, 76%); interrater reliability was specifically assessed 32% ($n = 8$) of the time. The pragmatism or feasibility of the tool was expressly assessed in about a quarter of the programs

Table 1 Types of practical-empirical support and fidelity domains assessed (*N* = 25)

Type of empirical support		Fidelity domains assessed	
Description	<i>N</i> (%)	Description	<i>N</i> (%)
Measure validity	9 (36%)	Adherence	24 (96%)
Interrater reliability	8 (32%)	Competence	5 (20%)
Pragmatism/feasibility	6 (24%)		
Measure reliability	5 (20%)		
Predictive validity	3 (12%)		
Association with outcomes	2 (8%)		
Variability and normality	1 (4%)		
Benchmarking	1 (4%)		
Comparative analysis	1 (4%)		
Racial disparities	1 (4%)		

Categories are not mutually exclusive

(*n* = 6, 24%). Several studies included less common determinations of the empirical basis of the fidelity approach. One study examined the extent to which fidelity measures identify racial disparities in treatment delivery outcomes (e.g., attrition; Yohannan et al., 2017). Two studies examined the relationship between fidelity ratings and treatment and clinical outcomes. One study (Breitenstein, Fogg, et al., 2010; Breitenstein, Gross, et al., 2010) found no association between fidelity and outcomes while the other study, focusing on organizational adherence (Brunk et al., 2014), found a strong relationship. Further, the vast majority of fidelity approaches focused on the domain of adherence (*n* = 24, 96%), with a smaller number examining competence (*n* = 5, 20%).

Strategies and Sources used to Determine Fidelity

Variable data collection strategies and informational sources supplied fidelity data (Table 2). The majority (*n* = 13; 52%) used multiple sources for obtaining fidelity data. The most common sources were observation (*n* = 13; 52%) and practitioner-reported fidelity checklists (*n* = 13; 52%), followed by interviews (*n* = 9; 36%). Less common was examination of service data (*n* = 5; 20%), reviews of clinical charts (*n* = 5; 20%), or review of client outcomes (*n* = 2; 8%).

Likewise, many articles obtained fidelity data from more than one informant source, which mirrors the aforementioned diversity in data collection strategies. The most common informant used was practitioner (*n* = 20; 80%), followed by administrators (*n* = 12; 48%), clients (*n* = 11; 44%), and supervisor or team leader (*n* = 4; 16%).

However, less variability was observed regarding who determined fidelity. Outside experts (e.g., trained reviewers, EBP purveyors) were largely responsible (*n* = 21; 84%) for determining fidelity. One study used computer software to

Table 2 Practical empirical approaches to fidelity (*N* = 25)

Data collection strategy	<i>n</i>	Data source	<i>n</i>	Fidelity arbitration	<i>n</i>	Practice or program	<i>n</i>	EBP Specific or agnostic/multiple	<i>n</i>	Level of fidelity	<i>n</i>
Multiple	13	Practitioners/teacher	20	Outside expert	21	Program	10	Specific	22	Team	12
Singular	12	Administrators	12	Practitioner or agency coders	3	Practice	10	Agnostic/multiple	3	Individual	12
Observation	13	Clients	11	Computer software	1	Both	5			Both	1
Fidelity checklists	13	Supervisors/ Team leaders	4								
Interviews	9										
Service data	5										
Clinical charts	5										
Client outcomes	2										

*Strategies and sources are not mutually exclusive

determine fidelity (Atkins et al., 2014). Another study determined fidelity by calculating inter-rater reliability between practitioners and independent coders (Breitenstein, Fogg, et al., 2010; Breitenstein, Gross, et al., 2010).

Fidelity Assessment Approaches

Program-level fidelity assessment includes those programs that assess the functioning of the entire EBP program, whereas practice level fidelity assessment focuses on “what happens in the room” between a practitioner and client. The number of articles that assessed fidelity at the program level ($n = 10$; 40%) and the practice level ($n = 10$; 40%) were equal. Five (20%) articles assessed fidelity at both program and practice levels.

When evaluating the extent to which fidelity approaches were specific to a single EBP (e.g., Multisystemic Therapy, family therapy for behavior problems) or were designed to be used across programs (i.e., EBP agnostic), the majority of approaches assessed fidelity specific to a single program ($n = 22$, 88%). However, fidelity approaches that were EBP agnostic were represented in three (12%) of the articles reviewed.

Some programs determined fidelity through examining aspects of team-level functioning while others determined fidelity at the individual practitioner level. There was an even split between the fidelity approaches that examined team-level fidelity ($n = 12$, 48%) compared with individual practitioner fidelity ($n = 12$, 48%). One program assessed fidelity at both levels (Brunk et al., 2014). Many of the articles reporting team-level approaches were reporting on fidelity to Assertive Community Treatment (ACT) (e.g., Bond et al., 2009a, 2009b; McGrew et al., 2011; Monroe-DeVita et al., 2011; Rollins et al., 2016, 2017). Of note, considerable overlap was observed in the practice vs. program and level of fidelity codes. All studies that assessed fidelity only at the practice level also assessed the level of fidelity only at the individual level. Nine of the ten studies that assessed only at the program level also assessed level of fidelity only at the team level.

Discussion

To date, no comprehensive review has specifically examined the empirical basis of fidelity strategies that can be administered within the settings in which the intervention was designed to be delivered (i.e., not strictly a research setting). Community-based agencies and practitioners delivering evidence-based interventions express interest in having practical approaches to fidelity monitoring and support (Kimber et al., 2019). The current review aimed to synthesize empirically supported practical approaches to fidelity measurement

in “real-world” settings. Having a better understanding of the current literature provides guidance to researchers and practitioners alike and identifies gaps and needs to encourage innovation for fidelity measurement.

Despite casting a broad net, only 25 articles met inclusionary criteria. Many of these articles focused on a limited number of interventions, the majority of which were clinically and programmatically intensive. These practical-empirical studies of fidelity provided evidence of the usefulness of the fidelity approach within naturalistic settings while specifically examining features of the fidelity strategy (e.g., the psychometric properties, feasibility, and uptake).

Examples in the review represented a broad range of strategies. Dorsey et al. (2017) shared an example of fidelity assessment that can be conducted during routine supervision sessions. Atkins et al. (2014) demonstrated that voice recognition technology provides an efficient assessment of video/audio-taped therapy sessions. Hogue et al. (2017) explained how to use benchmarking to facilitate assessment of fidelity. Essock et al. (2015) described how different types of data reported by different reporters provides program-level fidelity estimates. This emerging area of research represents an area of needed attention in the empirical literature.

Many articles were excluded from this review because the fidelity approach or tool was only developed for research studies and was not specifically assessed in a real-world setting. Such articles were determined to be *scientific-empirical* because they were tested only in a controlled research setting. There is potential for some of these approaches to become *practical-empirical* with future study. For example, Feely et al. (2018) outline a five-step process to develop a fidelity measurement system for research studies. Outlining such a process is very helpful in supporting the advancement of treatment *and* implementation research. However, to advance the prevalence of *practical-empirical* approaches, it is important to assess the practical relevance of the fidelity approach across each of these steps. This could involve asking additional questions such as: is the developed fidelity tool pragmatic for use in community-based settings? What resources does the tool require? Is there any specific training that would be necessary to administer the tool in community-based settings; and, is that training feasible or practical? How does the fidelity tool inform practice and program implementation?

On the other hand, the literature review identified multiple examples of practical approaches to fidelity measurement that lacked empirical basis (over a quarter of the articles extracted for review). In this case, the practical utility of the fidelity tool was demonstrated, but the manner and extent to which it functions to either predict practitioner adherence or competence in treatment delivery or its association with clinical outcomes remains unknown. Without this information, it is difficult to imagine how such fidelity support would be

useful to enhance practice. Yet this is consistent with other studies indicating that the vast majority of fidelity measurement approaches lack empirical support (Rolls-Reutz et al., 2020; Schoenwald & Garland, 2013). A recent analysis of 201 programs listed on the California Evidence-Based Clearinghouse's website (www.cebc4cw.org) revealed that approximately 30% of the programs that had a scientific rating of promising, supported, or well-supported had no fidelity assessment processes associated with their intervention (Rolls-Reutz et al., 2020). In this same study, the availability of empirical support for fidelity approaches was minimal (about 4%). Similar to the present study, Rolls-Reutz and colleagues found that there is substantial variability with regard to the methods and strategies used to determine fidelity within programs designed for community-based dissemination.

Expanding the list of practical and empirically supported fidelity approaches may be challenging. As mentioned, there are few researchers explicitly testing their fidelity approaches in this way. Moreover, the wide variability in approaches exemplifies that there is not a consistent approach that has broad acceptance in the scientific or practice communities. This review identified only twenty-five articles that described practical and empirically supported approaches to fidelity measurement. While meeting the definition of practical, the most common data collection strategy remained observation, which was used in individual and team-level approaches to fidelity assessment. Although this is considered a gold-standard strategy for fidelity assessment, it is perhaps the most labor intensive of the different strategies observed. Practitioner-reported checklists are similarly popular approaches, and are clearly more practical to implement. However, the correlation between checklists and observations is historically lower, though some more recent research indicates that there could be some support for checklists if they meet particular criteria (e.g., Hurlburt et al., 2010; Sheridan et al., 2009).

Combining observation or practitioner-report checklists with other common approaches to support fidelity could be a practical strategy to enhance fidelity measurement in community-based settings. In a review of cross-disciplinary training approaches to support transfer of training to practice, Lyon et al. (2011) suggest several strategies that are well aligned with fidelity support approaches (Lyon et al., 2011). For example, coaching is an effective strategy for supporting transfer of learning. Fidelity measurement could play a role in supporting “precision coaching,” ensuring that a practitioner receives feedback and supports that are specific to needed competencies (Weaver & DeRosier, 2019). These strategies, with some adjustments, could be viable at the team and individual practitioner level.

Finally, of note, emerging research indicates nearly inevitable adaptation when different communities implement

EBPs (Lau et al., 2017). This is an important area for further research and inquiry. Adaptations may address specific cultural groups (e.g., Baumann et al., 2015) or accommodate different inner and outer setting characteristics (Barnett et al., 2019; Damschroder et al., 2009). Stirman and colleagues developed a coding system for adaptations (Stirman et al., 2013). The extent to which and how agencies, clinical teams, and individual practitioners make adaptations to evidence-based programs and practices represents an important area of tension for fidelity measurement approaches (Forehand et al., 2010). None of the fidelity measurement approaches reviewed for this study expressly contained guidance about intervention flexibility. The nexus of these two areas of research is critical and will be further informed when intervention developers articulate the “active ingredients” in their interventions and implementation scientists provide enhanced guidance about how best to approach adaptations.

Limitations

Identifying articles for inclusion in this review was challenging. The term *practical-empirical* is not a standard search term. Thus, the review required substantial efforts to identify articles for inclusion. It is likely that studies were missed and findings from this review should be considered indicative of the state of the field, rather than summative. The operationalization of *practical-empirical* is not without critique. For this review, the operationalization of “practical” was given wide berth; articles meeting inclusionary criteria merely had to use the fidelity strategies in real-world settings. This purposely overly inclusive approach was taken because of the paucity of studies specifically looking at pragmatic approaches for fidelity monitoring. However, some may argue that approaches included in this review remain impractical for large-scale implementation.

Summary and Implications

This review summarizes and adds to the current state of the science as it relates to practical approaches to fidelity measurement. Increased emphasis on implementing EBPs with fidelity will lead agencies and practitioners to search for strategies to fulfill fidelity requirements. The results of this review indicate no consensus on a definitive approach that is practical and empirically supported. Because many EBPs lack standard fidelity structures (Rolls-Reutz et al., 2020), results from this review provide some insight for fidelity approaches and components that have an empirical basis and may be helpful to practitioners and agencies tasked with designing fidelity monitoring systems.

Several areas warrant further investigation. Few of the approaches described a comprehensive strategy of fidelity measurement; the most common element assessed was adherence. This greatly limits the opportunity to identify differential fidelity-related factors that could be associated with quality implementation and client outcomes. The over-focus on adherence may miss critical components of fidelity (e.g., competence) that could have substantial impacts on client experience. Further, the methods by which researchers evaluated their fidelity approaches varied substantially, and only a few examined the relationship with outcomes. With a few notable exceptions (e.g., Yohannan et al., 2017), researchers provided limited justification about the rationale for choosing the various strategies to determine empirical support for the fidelity tool or approach. In summary, there are few empirically validated approaches to determine fidelity that are used in real-world settings and most strategies rely on fidelity determinations by outside experts. This insufficiency represents a substantial implementation gap; unless the field identifies reliable approaches to fidelity supports that can be scaled, broad implementation of evidence-based treatments may be limited.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43477-021-00012-5>.

Declarations

Conflict of interest The authors report no conflict of interest.

References

- Atkins, D. C., Steyvers, M., Imel, Z. E., & Smyth, P. (2014). Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science*. <https://doi.org/10.1186/1748-5908-9-49>
- Barber, J. P., Triffleman, E., & Marmar, C. (2007). Considerations in treatment integrity: Implications and recommendations for PTSD research. *Journal of Traumatic Stress*, 20(5), 793–805
- Barnett, M. L., Brookman-Frazee, L., Gonzalez, J. C., Zhan, C., Rodriguez, A., Stadnick, N. A., & Lau, A. S. (2019). Qualitative reports of how and when therapists adapt children's evidence-based practices during community implementation. *Journal of Clinical Child & Adolescent Psychology*, 48(6), 894–905
- Barwick, M., Barac, R., Kimber, M., Akrong, L., Johnson, S. N., Cunningham, C. E., Bennett, K., Ashbourne, G., & Godden, T. (2020). Advancing implementation frameworks with a mixed methods case study in child behavioral health. *Translational Behavioral Medicine*, 10(3), 685–704
- Baumann, A. A., Powell, B. J., Kohl, P. L., Tabak, R. G., Penalba, V., Proctor, E. E., Domenech-Rodriguez, M. M., & Cabassa, L. J. (2015). Cultural adaptation and implementation of evidence-based parent-training: A systematic review and critique of guiding evidence. *Children and Youth Services Review*, 53, 113–120. <https://doi.org/10.1016/j.childyouth.2015.03.025>
- Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don't tell me: Behavioral rehearsal as a training and analogue fidelity tool. *Cognitive and Behavioral Practice*, 21(1), 1–11
- Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., Hurford, M. O., Hadley, T., Barg, F. K., Walsh, L. M., Adams, D. R., & Mandell, D. S. (2015). Predictors of community therapists' use of therapy techniques in a large public mental health system. *Journal of the American Medical Association Pediatrics*, 169(4), 374–382. <https://doi.org/10.1001/jamapediatrics.2014.3736>
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the national institute of health behavior change consortium. *Health Psychology*, 23(5), 443–451. <https://doi.org/10.1037/0278-6133.23.5.443>
- Bond, G. R., & Drake, R. E. (2020). Assessing the fidelity of evidence-based practices: History and current status of a standardized measurement methodology. *Administration and Policy in Mental Health and Mental Health Services Research*, 47, 874–884. <https://doi.org/10.1007/s10488-019-00991-6>
- Bond, G. R., Drake, R. E., McHugo, G. J., Rapp, C. A., & Whitley, R. (2009a). Strategies for improving fidelity in the national evidence-based practices project. *Research on Social Work Practice*, 19(5), 569–581. <https://doi.org/10.1177/1049731509335531>
- Bond, G. R., Drake, R. E., Rapp, C. A., McHugo, G. J., & Xie, H. (2009b). Individualization and quality improvement: Two new scales to complement measurement of program fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 36(5), 349–357. <https://doi.org/10.1007/s10488-009-0226-y>
- Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation research. *Mental Health Services Research*, 2(2), 75–87. <https://doi.org/10.1023/a:1010153020697>
- Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research*, 59(3), 158–165
- Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing and Health*, 33(2), 164–173
- Broderick, P. C., Frank, J. L., Berrena, E., Schussler, D. L., Kohler, K., Mitra, J., Khan, L., Levitan, J., Mahfouz, J., Shields, L., & Greenberg, M. T. (2019). Evaluating the quality of mindfulness instruction delivered in school settings: Development and validation of a teacher quality observational rating scale. *Mindfulness*, 10, 36–45. <https://doi.org/10.1007/s12671-018-0944-x>
- Brunk, M. A., Chapman, J. E., & Schoenwald, S. K. (2014). Defining and evaluating fidelity at the program level in psychosocial treatments: A preliminary investigation. *Zeitschrift für Psychologie*, 222(1), 22–29. <https://doi.org/10.1027/2151-2604/a000162>
- Bruns, E. J., Burchard, J. D., Suter, J. C., Leverenz-Brady, K., & Force, M. M. (2004). Assessing fidelity to a community-based treatment for youth: The wraparound fidelity index. *Journal of Emotional and Behavioral Disorders*, 12(2), 79–89. <https://doi.org/10.1177/10634266040120020201>
- Collyer, H., Eisler, I., & Woolgar, M. (2019). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child and Adolescent Psychiatry*, 29, 417–431
- Cuddeback, G. S., Morrissey, J. P., Domino, M. E., Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2013). Fidelity to recovery-oriented ACT practices and consumer outcomes. *Psychiatric Services*, 64(4), 318–323. <https://doi.org/10.1176/appi.ps.201200097>

- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(1), 50
- Dorsey, S., Lyon, A. R., Pullmann, M. D., Jungbluth, N., Berliner, L., & Beidas, R. (2017). Behavioral rehearsal for analogue fidelity: Feasibility in a state-funded children's mental health initiative. *Administration and Policy in Mental Health and Mental Health Services Research*, 44, 395–404. <https://doi.org/10.1007/s10488-016-0727-4>
- Dorsey, S., Pullmann, M. D., Deblinger, E., Berliner, L., Kerns, S. E., Thompson, K., Unutzer, J., Weisz, J., & Garland, A. F. (2013). Improving practice in community-based settings: A randomized trial of supervision-study protocol. *Implementation Science*. <https://doi.org/10.1186/1748-5908-8-89>
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Essock, S. M., Nossel, I. R., McNamara, K., Bennett, M. E., Buchanan, R. W., Kreyenbuhl, J. A., Pharm, D., Mendon, S. J., Goldman, H. H., & Dixon, L. B. (2015). Practical monitoring of treatment fidelity: Examples from a team-based intervention for people with early psychosis. *Psychiatric Services*, 66(7), 674–676. <https://doi.org/10.1176/appi.ps.201400531>
- Feely, M., Seay, K. D., Lanier, P., Auslander, W., & Kohl, P. L. (2018). Measuring fidelity in research studies: A field guide to developing a comprehensive fidelity measurement system. *Child and Adolescent Social Work Journal*, 35, 139–152. <https://doi.org/10.1007/s10560-017-0512-6>
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A synthesis of the literature*. Tampa Florida: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network (FMHI Publication #231).
- Forehand, R., Dorsey, S., Jones, D. J., Long, N., & McMahon, R. J. (2010). Adherence and flexibility: They can (and do) coexist! *Clinical Psychology: Science and Practice*, 17(3), 258–264
- Gillespie, M. L., Huey, S. J., Jr., & Cunningham, P. B. (2017). Predictive validity of an observer-rated adherence protocol for multisystemic therapy with juvenile drug offenders. *Journal of Substance Abuse Treatment*, 76, 1–10. <https://doi.org/10.1016/j.jsat.2017.01.001>
- Gingerich, S., & Mueser, K. T. (2010). *Illness management and recovery implementation resource kit, revised ed.* Rockville, Md, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.
- Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006). *MST therapist adherence measure-revised (TAM-R)*. Medical University of South Carolina, Family Services Research Center.
- Hoagwood, K. E. (2013). Don't mourn: Organize. Reviving mental health services research for healthcare quality improvement. *Clinical Psychology: Science and Practice*, 20(1), 120–126
- Hogue, A., Dauber, S., & Henderson, C. E. (2017). Benchmarking family therapy for adolescent behavior problems in usual care: Fidelity, outcomes, and therapist performance differences. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(5), 626–641. <https://doi.org/10.1007/s10488-016-0769-7>
- Hogue, A., Dauber, S., Henderson, C. E., & Liddle, H. A. (2014). Reliability of therapist self-report on treatment targets and focus in family-based intervention. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(5), 697–705. <https://doi.org/10.1007/s10488-013-0520-6>
- Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (2015). Validity of therapist self-report ratings of fidelity to evidence-based practices for adolescent behavior problems: Correspondence between therapists and observers. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(2), 229–243
- Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, 76(4), 544–555
- Hukkelberg, S. S., & Ogden, T. (2013). Working alliance and treatment fidelity as predictors of externalizing problem behaviors in parent management training. *Journal of Consulting and Clinical Psychology*, 81(6), 1010–1020. <https://doi.org/10.1037/a0033825>
- Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research*, 37(3), 230–244. <https://doi.org/10.1007/s10488-009-0251-x>
- Isett, K. R., Burnam, M. A., Coleman-Beattie, B., Hyde, P. S., Morrissey, J. P., Magnabosco, J., Rapp, C. A., Ganju, V., & Goldman, H. H. (2007). The state policy context of implementation issues for evidence-based practices in mental health. *Psychiatric Services*, 58(7), 914–921. <https://doi.org/10.1176/ps.2007.58.7.914>
- Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. *Professional Psychology: Research and Practice*, 38(1), 13–20
- Kimber, M., Barac, R., & Barwick, M. (2019). Monitoring fidelity to an evidence-based treatment: Practitioner perspectives. *Clinical Social Work Journal*, 47(2), 207–221
- Lau, A., Barnett, M., Stadnick, N., Saifan, D., Regan, J., Wiltsey Stirman, S., Roesch, S., & Brookman-Frazee, L. (2017). Therapist report of adaptations to delivery of evidence-based practices within a system-driven reform of publicly funded children's mental health services. *Journal of Consulting and Clinical Psychology*, 85(7), 664–675. <https://doi.org/10.1037/ccp0000215>
- Lyon, A. R., Stirman, S. W., Kerns, S. E. U., & Bruns, E. J. (2011). Developing the mental health workforce: Review and application of training approaches from multiple disciplines. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(4), 238–253. <https://doi.org/10.1007/s10488-010-0331-y>
- Margolies, P. J., Humensky, J. L., Chiang, I. C., Covell, N. H., Broadway-Wilson, K., Gregory, R., Jewell, T. C., Scannevin, G., Jr., Baker, S., & Dixon, L. B. (2017). Is there a role for fidelity self-assessment in the individual placement and support model of supported employment? *Psychiatric Services*, 68(9), 975–978. <https://doi.org/10.1176/appi.ps.201600264>
- Margolies, P. J., Humensky, J. L., Chiang, I. C., Covell, N. H., Jewell, T. C., Broadway-Wilson, K., Gregory, R., Scannevin, G., Jr., & Dixon, L. B. (2018). Relationship between self-assessed fidelity and self-reported employment in the individual placement and support model of supported employment. *Psychiatric Services*, 69(5), 609–612. <https://doi.org/10.1176/appi.ps.201700472>
- McGrew, J. H., White, L. M., Stull, L. G., & Wright-Berryman, J. (2011). A comparison of self-reported and phone-administered methods of ACT fidelity assessment: A pilot study in Indiana. *Psychiatric Services*, 64(3), 272–276. <https://doi.org/10.1176/appi.ps.001252012>
- McGuire, A. B., Luther, L., White, D., White, L. M., McGrew, J., & Salyers, M. P. (2016). The “critical” elements of illness management and recovery: Comparing methodological approaches.

- Administration and Policy in Mental Health and Mental Health Services Research*, 43(1), 1–10. <https://doi.org/10.1007/s10488-014-0614-9>
- McGuire, A. B., Stull, L. G., Mueser, K. T., Santos, M., Mook, A., Rose, N., Tunze, C., White, L. M., & Salyers, M. P. (2012). Development and reliability of a measure of clinician competence in providing illness management and recovery. *Psychiatric Services*, 63(8), 772–778. <https://doi.org/10.1176/appi.ps.201100144>
- McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2015). Measuring treatment differentiation for implementation research: The Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale. *Psychological Assessment*, 27(1), 314–325
- McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, 20(1), 14–32
- Melau, M., Albert, N., & Nordentoft, M. (2019). Programme fidelity of specialized early intervention in Denmark. *Early Intervention Psychiatry*, 13(3), 627–632. <https://doi.org/10.1111/eip.12549>
- Melde, C., Esbensen, F. A., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review*, 30(6), 714–740. <https://doi.org/10.1177/0193841X06293412>
- Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2011). The TMACT: A new tool for measuring fidelity to Assertive Community Treatment. *Journal of the American Psychiatric Nurses Association*, 17(1), 17–29. <https://doi.org/10.1177/1078390310394658>
- Morrison, J. D., Becker, H., & Stuijbergen, A. K. (2017). Evaluation of intervention fidelity in a multisite clinical trial in persons with Multiple Sclerosis. *Journal of Neuroscience Nursing*, 49(6), 344–348. <https://doi.org/10.1097/JNN.0000000000000315>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65–76
- Rollins, A. L., Kukla, M., Salyers, M. P., McGrew, J. H., Flanagan, M. E., Leslie, D. L., Hunt, M. G., & McGuire, A. B. (2017). Comparing the costs and acceptability of three fidelity assessment methods for Assertive Community Treatment. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(5), 810–816. <https://doi.org/10.1007/s10488-016-0785-7>
- Rollins, A. L., McGrew, J. H., Kukla, M., McGuire, A. B., Flanagan, M. E., Hunt, M. G., Leslie, D. L., Collins, L. A., Wright-Berryman, J. L., Hicks, L. J., & Salyers, M. P. (2016). Comparison of Assertive Community Treatment fidelity assessment methods: Reliability and validity. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(2), 157–167. <https://doi.org/10.1007/s10488-015-0641-1>
- Rolls-Reutz, J., Kerns, S. E. U., Sedivy, J. A., & Mitchell, C. (2020). Documenting the implementation gap, Part 1: Use of fidelity supports in programs indexed in the California Evidence-Based Clearinghouse. *Journal of Family Social Work*, 23, 114–132
- Sanches, S. A., Swildens, W. E., van Busschbach, J. T., Farkas, M. D., van Weeghel, J., & van Wel, T. (2018). FiRe: Evaluation of a fidelity measure to promote implementation of evidence-based rehabilitation. *Psychiatric Rehabilitation Journal*, 41(1), 46–54. <https://doi.org/10.1037/prj0000276>
- Schoenwald, S. K., & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment*, 25(1), 146–156. <https://doi.org/10.1037/a0029715>
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), 32–43
- Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research*, 5(4), 223–239
- Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review*, 38(4), 460–475
- Sedlar, G., Bruns, E. J., Walker, S. C., Kerns, S. E., & Negrete, A. (2017). Developing a quality assurance system for multiple evidence based practices in a statewide service improvement initiative. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(1), 29–41. <https://doi.org/10.1007/s10488-015-0663-8>
- Sheridan, S. M., Swanger-Gagné, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review*, 38(4), 476–495
- Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., Rodríguez, A., Cox, J. R., Reise, S. P., Bonifay, W. E., Weisz, J. R., & Kendall, P. C. (2016). Initial development of a treatment adherence measure for Cognitive-Behavioral Therapy for child anxiety. *Psychological Assessment*, 28(1), 70–80
- Stirman, S. W., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, 8, 65. <https://doi.org/10.1186/1748-5908-8-65>
- Suhrheinrich, J., Dickson, K. S., Chan, N., Chan, J. C., Wang, T., & Stahmer, A. C. (2019). Fidelity assessment in community programs: An approach to validating simplified methodology. *Behavior Analysis in Practice*, 13(1), 29–39. <https://doi.org/10.1007/s40617-019-00337-6>
- Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. C. (2012). Bridging research and practice: Models for dissemination and implementation research. *American Journal of Preventive Medicine*, 43(3), 337–350
- Teague, G. B., Mueser, K. T., & Rapp, C. A. (2012). Advances in fidelity measurement for mental health services research: Four measures. *Psychiatric Services*, 63(8), 765–771. <https://doi.org/10.1176/appi.ps.201100430>
- Walton, H., Spector, A., Tombor, I., & Michie, S. (2017). Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: A systematic review of measure quality. *British Journal of Health Psychology*, 22(4), 872–903
- Weaver, C., & DeRosier, M. E. (2019). Commentary on scaling-up evidence-based interventions in public systems. *Prevention Science*, 20(8), 1178–1188. <https://doi.org/10.1007/s11121-019-01059-5>
- Webb, C. A., Derubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200–211. <https://doi.org/10.1037/a0018912>
- Yohannan, J., Carlson, J. S., Shepherd, M., & Batsche-McKenzie, K. (2017). Exploring attrition, fidelity, and effectiveness of wrap-around services among low-income youth of different racial backgrounds. *Family, Systems, and Health*, 35(4), 430–438. <https://doi.org/10.1037/fsh0000313>