



# Measuring implementation outcomes: An updated systematic review of measures' psychometric properties

Implementation Research and Practice  
Volume 1: Jan-Dec 2020 1–29  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2633489520936644  
journals.sagepub.com/home/irp

Kayne Mettert<sup>1</sup>, Cara Lewis<sup>1</sup>, Caitlin Dorsey<sup>1</sup>, Heather Halko<sup>2</sup> and Bryan Weiner<sup>3</sup>

## Abstract

**Background:** Systematic reviews of measures can facilitate advances in implementation research and practice by locating reliable and valid measures and highlighting measurement gaps. Our team completed a systematic review of implementation outcome measures published in 2015 that indicated a severe measurement gap in the field. Now, we offer an update with this enhanced systematic review to identify and evaluate the psychometric properties of measures of eight implementation outcomes used in behavioral health care.

**Methods:** The systematic review methodology is described in detail in a previously published protocol paper and summarized here. The review proceeded in three phases. Phase I, data collection, involved search string generation, title and abstract screening, full text review, construct assignment, and measure forward searches. Phase II, data extraction, involved coding psychometric information. Phase III, data analysis, involved two trained specialists independently rating each measure using PAPERS (Psychometric And Pragmatic Evidence Rating Scales).

**Results:** Searches identified 150 outcomes measures of which 48 were deemed unsuitable for rating and thus excluded, leaving 102 measures for review. We identified measures of acceptability ( $N=32$ ), adoption ( $N=26$ ), appropriateness ( $N=6$ ), cost ( $N=31$ ), feasibility ( $N=18$ ), fidelity ( $N=18$ ), penetration ( $N=23$ ), and sustainability ( $N=14$ ). Information about internal consistency and norms were available for most measures (59%). Information about other psychometric properties was often not available. Ratings for internal consistency and norms ranged from “adequate” to “excellent.” Ratings for other psychometric properties ranged mostly from “poor” to “good.”

**Conclusion:** While measures of implementation outcomes used in behavioral health care (including mental health, substance use, and other addictive behaviors) are unevenly distributed and exhibit mostly unknown psychometric quality, the data reported in this article show an overall improvement in availability of psychometric information. This review identified a few promising measures, but targeted efforts are needed to systematically develop and test measures that are useful for both research and practice.

**Plain language abstract:** When implementing an evidence-based treatment into practice, it is important to assess several outcomes to gauge how effectively it is being implemented. Outcomes such as acceptability, feasibility, and appropriateness may offer insight into why providers do not adopt a new treatment. Similarly, outcomes such as fidelity and penetration may provide important context for why a new treatment did not achieve desired effects. It is important that methods to measure these outcomes are accurate and consistent. Without accurate and consistent measurement, high-quality evaluations cannot be conducted. This systematic review of published studies sought to identify questionnaires (referred to as measures) that ask staff at various levels (e.g., providers, supervisors) questions

<sup>1</sup>Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

<sup>2</sup>Judge Baker Children's Center, Harvard Medical School, Boston, MA, USA

<sup>3</sup>Department of Global Health, University of Washington, Seattle, WA, USA

## Corresponding author:

Kayne Mettert, Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101-1466, USA.  
Email: kayne.d.metttert@kp.org



related to implementation outcomes, and to evaluate the quality of these measures. We identified 150 measures and rated the quality of their evidence with the goal of recommending the best measures for future use. Our findings suggest that a great deal of work is needed to generate evidence for existing measures or build new measures to achieve confidence in our implementation evaluations.

## Keywords

Implementation outcomes, implementation, dissemination, evidence based, behavioral health, measurement evaluation

It is well established that evidence-based practices are slow to be implemented into routine care (Carnine, 1997). Implementation science seeks to narrow the research-to-practice gap by identifying barriers and facilitators to effective implementation and designing strategies to achieve desired implementation outcomes. Proctor and colleagues' seminal work (Proctor et al., 2009, 2011) articulated at least eight implementation outcomes for the field: (1) *acceptability*, (2) *adoption*, (3) *appropriateness*, (4) *cost*, (5) *feasibility*, (6) *fidelity*, (7) *penetration*, and (8) *sustainability* (Table 1) (Lewis et al., 2015). While many other implementation frameworks exist (Glasgow et al., 1999), the outcomes framework developed by Proctor and colleagues provides clear differentiation of implementation outcomes from clinical and services outcomes and offers a harmonized focus for the field. Despite this, measurement inherently lags behind framework development and construct specification. Indeed, a 2015 systematic review led by our team found 104 measures used in behavioral health (including mental health, substance use, and other addictive behaviors) across these eight outcomes, with 50 identified for *acceptability*, 19 for *adoption*, but fewer than 10 for each of the other six outcomes. Our systematic review and psychometric assessment revealed that evidence of measures' reliability and validity is largely unknown or poor (Lewis et al., 2015). Of the psychometric properties assessed, only four measures had been tested for responsiveness or sensitivity to change, meaning that it is not clear whether the majority of implementation outcome measures are designed and able to detect change over time. Without accurate measurement of implementation outcomes, we cannot be sure if implementation efforts are (un)successful or if the measures are simply unfit to identify change in outcomes when it in fact occurs.

Since 2015, we have sought to update and expand these reviews. Full details about our updated approach are published in a protocol paper (Lewis et al., 2018). Three major differences are worth noting. One, we expanded our assessment of measures to their scales given that many measures purportedly assess numerous constructs delineated by scales. For example, the Texas Christian University Organizational Readiness for Change Scale contains 19 scales measuring constructs such as *motivation for change*, *available resources*, *staff attributes*, and *organizational climate* (Lehman et al., 2002). Moreover, researchers tend

to select and deploy individual scales versus full measures given their interest in minimizing respondent burden. Two, we added additional validity assessments including convergent validity, discriminant validity, concurrent validity, and known-groups validity. Three, we used a more comprehensive and critical approach to mapping measure content (i.e., items) to constructs by engaging content experts in the review of measure items: If at least two items in a scale reflected the definition of a given construct, an assignment was made in an effort to reflect real-world practices of measure use. Taken together, these changes will offer much richer and more useful information to implementation researchers and practitioners when selecting measures.

Although it has only been 4 years since the publication of our initial systematic review (Lewis et al., 2015), the field of implementation science has evolved with rapid pace. This progress, together with the enhancements made to our systematic review protocol (Lewis et al., 2018), calls for an update to our assessment of measures of implementation outcomes. Specifically, this article presents the findings from systematic reviews of the eight implementation outcomes, including a robust synthesis of psychometric evidence for all identified measures.

## Method

### Design overview

The systematic literature review and synthesis consisted of three phases. Phase I, measure identification, included the following five steps: (1) search string generation, (2) title and abstract screening, (3) full text review, (4) measure assignment to implementation outcome(s), and (5) measure forward (cited-by) searches. Phase II, data extraction, consisted of coding relevant psychometric information, and in Phase III data analysis was completed.

### Phase I: data collection

First, literature searches were conducted in PubMed and Embase bibliographic databases using search strings curated in consultation from PubMed support specialists and a library scientist. Consistent with our funding source and aim to identify and assess implementation-related measures in mental and behavioral health, our search was

built on four core levels: (1) terms for implementation (e.g., diffusion, knowledge translation, adoption); (2) terms for measurement (e.g., instrument, survey, questionnaire); (3) terms for evidence-based practice (e.g., innovation, guideline, empirically supported treatment); and (4) terms for behavioral health (e.g., behavioral medicine, mental disease, psychiatry) (Lewis et al., 2018). For the current study, we included a fifth level for each of the following *Implementation Outcomes* from Proctor et al. (2011): (1) *acceptability*, (2) *adoption*, (3) *appropriateness*, (4) *cost*, (5) *feasibility*, (6) *fidelity*, (7) *penetration*, and (8) *sustainability*. Literature searches were conducted independently for each outcome, thus eight different sets of search strings were employed. Articles published from 1985 to 2017 were included in the search. Searches were completed from April 2017 to May 2017 (Table 2).

Identified articles were vetted through a title and abstract screening followed by full text review to confirm relevance to the study parameters. In brief, we included empirical studies and measure development studies that contained one or more quantitative measures of any of the eight implementation outcomes if they were used in an evaluation of an implementation effort in a behavioral health context. Of note, we decided to retain only *fidelity* measures that were not specific to one evidence-based practice (EBP) and could be applied generally to be consistent with our goal of identifying broadly applicable measures.

Included articles then progressed to the fourth step, construct assignment. Trained research specialists (C.D., K.M.) mapped measures and/or their scales to one or more of the eight aforementioned implementation outcomes (Proctor et al., 2011). Assignment was based on the study author's definition of what was being measured. Assignment was also based on content coding by the research team who reviewed all items of the measure for evidence of content explicitly assessing one of the eight implementation outcomes when compared against the construct definition. Construct assignment was checked and confirmed by content expert (C.L.) having reviewed items within each measure and/or scale.

The final step subjected the included measures to "cited-by" searches in PubMed and Embase to identify all empirical articles that used the measure in behavioral health implementation research.

### *Phase II: data extraction*

Once all relevant literature was retrieved, articles were compiled into "measure packets." These measure packets included the measure itself (as available), the measurement development article(s) (or article with the first empirical use in a behavioral health context), and all additional empirical uses of the measure in behavioral health. In order to identify all relevant reports of

psychometric information, the team of trained research specialists (CD, KM) reviewed each article and electronically extracted information to assess the psychometric and pragmatic rating criteria, referred to hereafter as PAPERS (Psychometric And Pragmatic Evidence Rating Scale). The full rating system and criteria for the PAPERS are published elsewhere (Lewis et al., 2018; Stanick et al., 2019). The current study, which focuses on psychometric properties only, used nine relevant PAPERS criteria: (1) internal consistency, (2) convergent validity, (3) discriminant validity, (4) known-groups validity, (5) predictive validity, (6) concurrent validity, (7) structural validity, (8) responsiveness, and (9) norms. Data on each psychometric criterion were extracted for both full measure and individual scales as appropriate. Measures were considered "unsuitable for rating" if the format of construct assessment did not produce psychometric information (e.g., qualitative nomination form) or format of the measure did not conform to the rating scale (e.g., cost analysis formula, penetration formula).

Having extracted all data related to psychometric properties, the quality of information for each of the nine criteria was rated using the following scale: "poor" (-1), "none" (0), "minimal/emerging" (1), "adequate" (2), "good" (3), or "excellent" (4). Final ratings were determined from either a single score or a "rolled up median" approach. If a measure was unidimensional or the measure had only one rating for a criterion in an article packet, then this value was used as the final rating and no further calculations were conducted. If a measure had multiple ratings for a criterion across several articles in a packet, we calculated the median score across articles to generate the final rating for that measure on that criterion. For example, if a measure was used in four different studies, each of which rated internal consistency, we calculated the median score across all four articles to determine the final rating of internal consistency for that measure. This process was conducted for each criterion.

If a measure contained a subset of scales relevant to a construct, the ratings for those individual scales were "rolled up" by calculating the median which was then assigned as the final aggregate rating for the whole measure. For example, if a measure had four scales relevant to *acceptability* and each was rated for internal consistency, the median of those ratings was calculated and assigned as the final rating of internal consistency for that whole measure. This process was carried out for each psychometric criterion. When reporting the "rolled up median" approach, if the computed median resulted in a non-integer rating, the non-integer was rounded down (e.g., internal consistency ratings of 2 and 3 would result in a 2.5 median which was rounded down to 2). In cases where the median of two scores would equal "0" (e.g., a score of -1 and 1), the lower score would be taken (e.g., -1).

In addition to psychometric data, descriptive data were also extracted on each measure. Characteristics included

(1) country of origin, (2) concept defined by authors, (3) number of articles contained in each measure packet, (4) number of scales, (5) number of items, (6) setting in which measure had been used, (7) level of analysis, (8) target problem, and (9) stage of implementation as defined by the Exploration, Adoption/Preparation, Implementation, Sustainment (EPIS) model (Aarons et al., 2011).

### Phase III: data analysis

Simple statistics (i.e., frequencies) were calculated to report on measure characteristics and availability of psychometric-relevant data. A total score was calculated for each measure by summing the scores given to each of the nine psychometric criteria. The maximum possible rating for a measure was 36 (i.e., each criterion rated 4) and the minimum was -9 (i.e., each criterion rated -1). Bar charts were generated to display visual head-to-head comparisons across all measures within a given construct.

## Results

Following the rolled-up approach applied in this study, results are presented at the full measure level. Where appropriate, we indicate the number of scales relevant to a construct within that measure (see Figures A1–A8 in the Appendix 1 for PRISMA flowcharts of included and excluded studies).

### Overview of measures

Searches of electronic databases yielded 150 measures related to the eight implementation outcomes (*acceptability, adoption, appropriateness, cost, feasibility, fidelity, penetration, and sustainability*) that have been used in mental or behavioral health care research. Thirty-two measures of *acceptability* were identified, one of which was a specific scale within a broader measure (i.e., SFTRC Course Evaluation—Attitude scale) (Haug et al., 2008). Twenty-six measures of *adoption* were identified, three of which were scales part of broader measures (e.g., Perceptions of Computerized Therapy Questionnaire—Future Use Intentions scale) (Carper et al., 2013) and two of which were deemed “unsuitable for rating.” As mentioned previously, measures were considered “unsuitable for rating” if the format of construct assessment did not produce psychometric information or format of the measure did not conform to the rating scale (e.g., Fortney Measure of Adoption Rate) (Fortney et al., 2012). Six measures of *appropriateness* were identified, of which one was a scale within a broader measure (i.e., Moore & Benbasat Adoption of IT Innovation Measure—Compatibility scale) (Moore & Benbasat, 1991). Eighteen measures of *feasibility* were identified, four of which were scales within broader measures (e.g., Behavioral

Interventionist Satisfaction Survey—Feasibility scale) (McLean, 2013). Eighteen measures of *fidelity* were identified. Twenty-three measures of *penetration* were identified, of which 14 were deemed unsuitable for rating (e.g., Pace Proportion Measure of Penetration) (Pace et al., 2014). Finally, 14 measures of *sustainability* were identified, one of which was deemed unsuitable for rating (i.e., Kirchner Sustainability Measure) (Kirchner et al., 2014), and another was a scale within a larger measure (i.e., Eisen Provider Knowledge & Attitudes Survey—Sustainability scale) (Eisen et al., 2013). Thirty-one measures of implementation *cost* were identified; however, none of them were suitable for rating and thus their psychometric evidence was not assessed. It is worth noting that the number of measures listed above for each outcome does not add up to 150. This is because there were 14 measures identified that had scales relevant in multiple different outcomes. Of these 14 measures, 11 were included in two outcomes, one was included in three outcomes, and one was included in four.

### Characteristics of measures

Table 3 presents the descriptive characteristics of measures used to assess implementation outcomes. Most measures of implementation outcomes that were suitable for rating were used only once ( $n=78$ , 79%) and most were created in the United States ( $n=85$ , 86%). The remaining measures were developed in Australia, Canada, the United Kingdom, Israel, the Netherlands, Spain, and Zimbabwe. The majority of identified measures were used in the outpatient community setting ( $n=53$ , 54%) or a variety of “other” settings (e.g., prison, church) ( $n=41$ , 42%). Half of the measures were used to assess implementation outcomes influencing implementation in the general mental health field or substance use ( $n=49$ , 50%) and were applied at the implementation or sustainment stage ( $n=64$ , 65%, respectively). A small number of measures showed evidence of predictive validity for other implementation outcomes ( $n=15$ , 15%). Of these, six predicted *fidelity* (6%), five predicted *sustainability* (5%), two predicted *adoption* (2%), and two predicted *penetration* (2%).

### Availability of psychometric evidence

Of the 150 measures of implementation constructs, 48 were categorized as unsuitable for rating; unsurprisingly the majority of which were cost measures ( $n=31$ ). For the remaining 102 measures, there was limited psychometric information available (Table 4). Forty-six (45%) measures had no information for internal consistency, 80 (78%) had no information for convergent validity, 97 (94%) had no information for discriminant validity, 93 (91%) had no information for concurrent validity, 81 (80%) had no information for predictive validity, 88 (86%) had no

**Table 1.** Definitions of implementation outcomes.

Acceptability	The perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory.
Adoption	The intention, initial decision, or action to try or employ an innovation or evidence-based practice.
Appropriateness	The perceived fit, relevance, or compatibility of the innovation or evidence-based practice for a given practice setting, provider, or consumer; and/or perceived fit of the innovation to address a particular issue or problem.
Cost	The cost impact of an implementation effort.
Feasibility	The extent to which a new treatment, or an innovation, can be successfully used or carried out within a given agency or setting.
Fidelity	The degree to which an intervention was implemented as it was prescribed in the original protocol or as it was intended by the program developers.
Penetration	The integration of a practice within a service setting and its subsystems.
Sustainability	The extent to which a newly implemented treatment is maintained or institutionalized within a service setting's ongoing, stable operation.

Source: From Proctor et al. (2009, 2011).

information for known-groups validity, 84 (82%) had no information for structural validity, 95 (93%) had no information for responsiveness, and finally, 46 (45%) had no information on norms.

### Psychometric evidence rating scale results

Table 5 describes the median ratings and range of ratings for psychometric properties for measures deemed suitable for rating ( $n=102$ ) and those for which information was available (i.e., those with non-zero ratings on PAPERS criteria). Individual ratings for all measures can be found in Table 6 and head-to-head bar graphs can be found in Figures 1 to 7.

**Acceptability.** Thirty-two measures of *acceptability* were identified in mental or behavioral health care research. Information about internal consistency was available for 19 measures, convergent validity for 10 measures, discriminant validity for no measures, concurrent validity for one measure, predictive validity for six measures, known-groups validity for four measures, structural validity for four measures, responsiveness for two measures, and norms for twenty measures. For those measures of *acceptability* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “3—good,” “2—adequate” for convergent validity, “2—adequate” for concurrent validity, “-1—poor” for predictive validity, “-1—poor” for known-groups validity, “-1—poor” for structural validity, “1—minimal/emerging” for responsiveness, and “1—minimal/emerging” for norms.

The Pre-referral Intervention Team Inventory had the highest psychometric rating score among measures of *acceptability* used in mental and behavioral health care (psychometric total maximum score=14; maximum possible score=36), with ratings of “4—excellent” for internal consistency, “3—good” for convergent validity, “4—excellent” for structural validity, and “3—good” for

norms (Yetter, 2010). There was no information available on any of the remaining psychometric criteria.

**Adoption.** Twenty-six measures of *adoption* were identified in mental or behavioral health care research, two of which were deemed unsuitable for rating. Information about internal consistency was available for 16 measures, convergent validity for four measures, discriminant validity for no measures, concurrent validity for two measures, predictive validity for five measures, known-groups validity for two measures, structural validity for four measures, responsiveness for one measure, and norms for 12 measures. For those measures of *adoption* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,” “2—adequate” for convergent validity, “2—adequate” for concurrent validity, “1—minimal/emerging” for predictive validity, “-1—poor” for known-groups validity, “1—minimal/emerging” for structural validity, “1—minimal/emerging” for responsiveness, and “2—adequate” for norms.

The Williams “Intention to Adopt” and Ruzek “Measure of Adoption” measures had the highest psychometric rating scores among measures of *adoption* used in mental and behavioral health care (psychometric total maximum score=9; maximum possible score=36), with ratings of “3—good” and “2—adequate” for internal consistency, “3—good” and “0—no evidence” for convergent validity, “0—no evidence” and “3—good” for predictive validity, and “3—good” and “4—excellent” for norms, respectively (Ruzek et al., 2016; Williams et al., 2017). There was no information available on any of the remaining psychometric criteria. It is worth noting that these scores reflect the median of over 88 uses of this measure in behavioral health research.

**Appropriateness.** Six measures of *appropriateness* were identified in mental or behavioral health research. Information about internal consistency was available for three

**Table 2.** Database search terms.

Search term	Search string
Implementation	(Adopt[tiab] OR adopts[tiab] OR adopted[tiab] OR adoption[tiab] NOT “adoption”[MeSH Terms] OR Implement[tiab] OR implements[tiab] OR implementation[tiab] OR implementation[ot] OR “health plan implementation”[MeSH Terms] OR “quality improvement*”[tiab] OR “quality improvement”[tiab] OR “quality improvement”[MeSH Terms] OR diffused[tiab] OR diffusion[tiab] OR “diffusion of innovation”[MeSH Terms] OR “health information exchange”[MeSH Terms] OR “knowledge translation*”[tw] OR “knowledge exchange*”[tw]) AND
Evidence-based practice	(“empirically supported treatment”[All Fields] OR “evidence based practice*”[All Fields] OR “evidence based treatment”[All Fields] OR “evidence-based practice”[MeSH Terms] OR “evidence-based medicine”[MeSH Terms] OR innovation[tw] OR guideline[pt] OR (guideline[tiab] OR guideline’[tiab] OR guideline”[tiab] OR guideline’pregnancy[tiab] OR guideline’s[tiab] OR guideline l [tiab] OR guideline2015[tiab] OR guidelinebased[tiab] OR guidelineed[tiab] OR guidelinedevelopment[tiab] OR guidelinee[tiab] OR guidelineitem[tiab] OR guidelineon[tiab] OR guideliner[tiab] OR guideliner’[tiab] OR guidelinerecommended[tiab] OR guidelinerelated[tiab] OR guidelinertrade[tiab] OR guidelines[tiab] OR guidelines’[tiab] OR guidelines’quality[tiab] OR guidelines’s[tiab] OR guidelines l [tiab] OR guidelines l9[tiab] OR guidelines2[tiab] OR guidelines20[tiab] OR guidelinesfemale[tiab] OR guidelinesfor[tiab] OR guidelinesin[tiab] OR guidelinesmay[tiab] OR guidelineson[tiab] OR guideliness[tiab] OR guidelinesthat[tiab] OR guidelinestrade[tiab] OR guidelineswiki[tiab]) OR “guidelines as topic”[MeSH Terms] OR “best practice*”[tw]) AND
Measure	(instrument[tw] OR (survey[tw] OR survey’[tw] OR survey’s[tw] OR survey l00[tw] OR survey l2[tw] OR survey l988[tw] OR survey226[tw] OR survey36[tw] OR surveyability[tw] OR surveyable[tw] OR surveyance[tw] OR surveyans[tw] OR surveyansin[tw] OR surveybetween[tw] OR surveyd[tw] OR surveydagger[tw] OR surveydata[tw] OR surveydelhi[tw] OR surveyed[tw] OR surveyedandtestedthe[tw] OR surveyedpopulation[tw] OR surveyees[tw] OR surveyelicited[tw] OR surveyer[tw] OR surveyes[tw] OR surveyeyed[tw] OR surveyform[tw] OR surveyfreq[tw] OR surveygizmo[tw] OR surveyin[tw] OR surveying[tw] OR surveying’[tw] OR surveyings[tw] OR surveylogistic[tw] OR surveymaster[tw] OR surveymeans[tw] OR surveymeter[tw] OR surveymonkey[tw] OR surveymonkey’s[tw] OR surveymonkeytrade[tw] OR surveyng[tw] OR surveyor[tw] OR surveyor’[tw] OR surveyor’s[tw] OR surveyors[tw] OR surveyors’[tw] OR surveyortrade[tw] OR surveypatients[tw] OR surveyphreg[tw] OR surveyplus[tw] OR surveyprocess[tw] OR surveyreg[tw] OR surveys[tw] OR surveys’[tw] OR surveys’food[tw] OR surveys’usefulness[tw] OR surveysclub[tw] OR surveyselect[tw] OR surveyset[tw] OR surveyset’[tw] OR surveyspot[tw] OR surveystrade[tw] OR surveysuite[tw] OR surveytaken[tw] OR surveythese[tw] OR surveytm[tw] OR surveytracker[tw] OR surveytrade[tw] OR surveyvas[tw] OR surveywas[tw] OR surveywiz[tw] OR surveyxact[tw]) OR (questionnaire[tw] OR questionnaire’[tw] OR questionnaire’07[tw] OR questionnaire’midwife[tw] OR questionnaire’s[tw] OR questionnaire l [tw] OR questionnaire l l [tw] OR questionnaire l2[tw] OR questionnaire2[tw] OR questionnaire25[tw] OR questionnaire3[tw] OR questionnaire30[tw] OR questionnaireand[tw] OR questionnairebased[tw] OR questionnairebefore[tw] OR questionnaireconsisted[tw] OR questionnairecopyright[tw] OR questionnaired[tw] OR questionnairedeveloped[tw] OR questionnaireepq[tw] OR questionnaireforpediatric[tw] OR questionnairegrtr[tw] OR questionnairehas[tw] OR questionnaireitaq[tw] OR questionnairel02[tw] OR questionnairemcscscale[tw] OR questionnaireurse[tw] OR questionnaireon[tw] OR questionnaireonline[tw] OR questionnairepf[tw] OR questionnairephq[tw] OR questionnaires[tw] OR questionnaires’[tw] OR questionnaires”[tw] OR questionnairescan[tw] OR questionnairesdql l adolescent [tw] OR questionnaireess[tw] OR questionnairetrade[tw] OR questionnaireeure[tw] OR questionnaireev[tw] OR questionnairewere[tw] OR questionnairex[tw] OR questionnairey[tw]) OR instruments[tw] OR “surveys and questionnaires”[MeSH Terms] OR “surveys and questionnaires”[MeSH Terms] OR measure[tiab] OR (measurement[tiab] OR measurement’[tiab] OR measurement’s[tiab] OR measurement l [tiab] OR measuremental[tiab] OR measurementd[tiab] OR measuremented[tiab] OR measurementexhaled[tiab] OR measurementff[tiab] OR measurementin[tiab] OR measuremention[tiab] OR measurementis[tiab] OR measurementkcomputation[tiab] OR measurementl[tiab] OR measurementmanometry[tiab] OR measurementmethods[tiab] OR measurementof[tiab] OR measurementon[tiab] OR measurementpro[tiab] OR measurementresults[tiab] OR measurements[tiab] OR measurements’[tiab] OR measurements’s[tiab] OR measurements0[tiab] OR measurements5[tiab] OR measurementsa[tiab] OR measurementsare[tiab] OR measurementscanbe[tiab] OR measurementscheme[tiab] OR measurementsfor[tiab] OR measurementsgave[tiab] OR measurementsin[tiab] OR measurementsindicate[tiab] OR measurementsmoking[tiab] OR measurementsof[tiab] OR measurementson[tiab] OR measurementsreveal[tiab] OR measurementss[tiab] OR measurementswere[tiab] OR measurementtime[tiab] OR measurements[tiab] OR measurementusing[tiab] OR measurementws[tiab]) OR measures[tiab] OR inventory[tiab]) AND

(Continued)

**Table 2.** (Continued)

Search term	Search string
Mental health	("mental health"[tw] OR "behavioral health"[tw] OR "behavioural health"[tw] OR "mental disorders"[MeSH Terms] OR "psychiatry"[MeSH Terms] OR psychiatry[tw] OR psychiatric[tw] OR "behavioral medicine"[MeSH Terms] OR "mental health services"[MeSH Terms] OR (psychiatrist[tw] OR psychiatrist'[tw] OR psychiatrist's[tw] OR psychiatrists[tw] OR psychiatrists'[tw] OR psychiatrists'awareness[tw] OR psychiatrists'opinion[tw] OR psychiatrists'quality[tw] OR psychiatristsand[tw] OR psychiatristsare[tw]) OR "hospitals, psychiatric"[MeSH Terms] OR "psychiatric nursing"[MeSH Terms]) AND
Acceptability	acceptability[tw] OR satisfaction [tw] OR agreeable[tw] OR
Adoption	(adopt[tw] OR adopts[tw] OR adopted[tw] OR adoption[tw] OR (("intention"[MeSH Terms] OR "intention"[All Fields]) AND adopt[All Fields]) NOT "adoption"[MeSH Terms] OR uptake[tw] OR utilization[tw] OR "initial implementation"[All Fields]) OR
Appropriateness	appropriateness[tw] OR applicability[tw] OR applicability[tw] OR compatibility[tw] OR "perceived fit" OR fitness[tw] OR sustainability[tw] OR relevance[tw] OR relevant[tw] OR suitability[tw] OR usefulness[tw] OR practicability[tw] OR
Cost	"marginal cost" [tw] OR "cost-effectiveness" OR economics[sh] OR economics[mh] OR cost-benefit analysis[mh] OR "cost benefit"[tw] OR "cost utility" OR
Feasibility	feasibility[tw] OR transferability [tw] OR applicability [tw] OR practicability [tw] OR workability [tw] OR "actual fit" [tw] OR "actual utility" [tw] OR "suitability for everyday use"[tw] OR
Fidelity	fidelity [tw] OR "delivered as intended" OR adherence[tw] OR "patient adherence"[tw] OR patient compliance[mh] OR compliance[tw] OR compliant[tw] OR integrity [tw] OR "quality of program delivery"[tw] OR
Penetration	penetration [tw] OR "integration of practice"[tw] OR infiltration [tw] OR
Sustainability	sustain*[tw] OR maintenance[tw] OR "long-term implementation" OR routinization [tw] OR durability [tw] OR institutionalization [tw] OR "capacity building" [tw] OR continuation [tw] OR incorporation [tw] OR integration [tw] OR "sustained use"[tw]

measures, convergent validity for one measure, discriminant validity for no measures, concurrent validity no measures, predictive validity for one measure, known-groups validity for no measures, structural validity for one measure, responsiveness for no measures, and norms for two measures. For those measures of *appropriateness* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was "4—excellent," "3—good" for convergent validity, "3—good" for predictive validity, "4—excellent" for structural validity, and "1—minimal/emerging" for norms.

The Pre-referral Intervention Team Inventory had the highest psychometric rating score among measures of *appropriateness* used in mental and behavioral health care (psychometric total maximum score=14; maximum possible score=36), with ratings of "4—excellent" for internal consistency, "3—good" for convergent validity, "4—excellent" for structural validity, and "3—good" for norms (Yetter, 2010). There was no information available on any of the remaining psychometric criteria.

**Cost.** Thirty-one measures of implementation *cost* were identified in mental or behavioral health research;

however, none of them were suitable for rating and thus their psychometric information was not assessed.

**Feasibility.** Eighteen measures of *feasibility* were identified in mental or behavioral health research. Information about internal consistency was available for five measures, convergent validity for no measures, discriminant validity for no measures, concurrent validity no measures, predictive validity for one measure, known-groups validity for one measure, structural validity for one measure, responsiveness for no measures, and norms for nine measures. For those measures of *feasibility* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was "4—excellent," "1—minimal/emerging" for predictive validity, "1—poor" for known-groups validity, "2—adequate" for structural validity, and "1—poor" for norms.

The Children's Usage Rating Profile had the highest psychometric rating score among measures of *feasibility* used in mental and behavioral health care (psychometric total maximum score=6; maximum possible score=36), with ratings of "4—excellent" for internal consistency and "2—adequate" for norms (Briesch & Chafouleas, 2009).

**Table 3.** Description of measures and subscales.

	Acceptability (N=32)		Adoption (N=25)		Appropriateness (N=6)		Feasibility (N=18)		Fidelity (N=18)		Penetration (N=9)		Sustainability (N=13)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Concept defined														
Yes	6	19	10	40	5	83	2	11	5	28	9	100	2	15
No	26	81	15	60	1	17	16	89	13	72	0	0	11	85
One-time use only														
Yes	14	44	18	72	5	83	10	56	11	61	5	56	7	54
No	18	56	7	28	1	17	8	44	7	39	4	44	6	46
Number of items														
1–5	1	3	4	16	0	0	2	11	2	11	1	11	0	0
6–10	4	13	2	8	0	0	4	22	0	0	0	0	0	0
11 or more	25	78	16	64	5	83	10	56	11	61	5	56	11	85
Not specified	2	6	3	12	1	17	2	11	5	28	3	33	2	15
Country														
US	21	66	15	60	3	50	9	50	15	83	7	78	12	92
Other	11	34	10	40	3	50	9	50	3	17	2	22	1	8
Setting														
State mental health	1	3	1	4	0	0	0	0	0	0	0	0	0	0
Inpatient psychiatry	1	3	1	4	0	0	1	6	1	6	0	0	0	0
Outpatient community	11	34	12	48	1	17	7	39	8	44	4	44	6	46
School mental health	15	47	3	12	1	17	1	6	1	6	0	0	1	8
Residential care	1	3	4	16	0	0	0	0	2	11	0	0	0	0
Other	4	13	7	28	1	17	4	22	6	33	4	44	4	31
Not specified	0	0	0	0	3	50	4	22	1	6	1	11	2	15
Level														
Consumer	5	16	1	4	1	17	2	11	0	0	1	11	0	0
Organization	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Clinic/site	0	0	0	0	0	0	0	0	0	0	1	11	1	8
Provider	20	63	19	76	0	0	13	72	15	83	7	78	9	69
System	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Team	0	0	0	0	0	0	0	0	1	6	0	0	0	0
Director	1	3	3	12	0	0	0	0	0	0	0	0	1	8
Supervisor	3	9	1	4	1	17	1	6	1	6	1	11	1	8
Other	2	6	4	16	1	17	1	6	1	6	0	0	1	8
Not specified	2	6	2	8	3	50	1	6	0	0	0	0	2	15
Population														
General mental health	3	9	6	24	0	0	2	11	6	33	5	56	3	23
Anxiety	2	6	1	4	0	0	0	0	1	6	0	0	0	0
Depression	4	13	2	8	0	0	2	11	3	17	1	11	3	23
Suicidal ideation	1	3	0	0	0	0	0	0	1	6	1	11	0	0
Alcohol use disorder	0	0	1	4	0	0	0	0	0	0	0	0	0	0
Substance use disorder	5	16	10	40	1	17	5	28	5	28	0	0	3	23
Behavioral disorder	12	38	1	4	0	0	1	6	0	0	0	0	1	8
Mania	1	3	0	0	0	0	2	11	0	0	0	0	0	0
Eating disorder	3	9	0	0	0	0	0	0	0	0	0	0	0	0
Grief	1	3	0	0	0	0	0	0	0	0	0	0	0	0
Tic disorder	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trauma	1	3	3	12	0	0	0	0	2	11	0	0	1	8
Other	6	19	5	20	1	17	3	17	1	6	0	0	1	8
Not specified	0	0	0	0	4	67	2	11	1	6	2	22	1	8
EPIS phase														
Exploration	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Preparation	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Continued)



**Table 3.** (Continued)

	Acceptability (N=32)		Adoption (N=25)		Appropriateness (N=6)		Feasibility (N=18)		Fidelity (N=18)		Penetration (N=9)		Sustainability (N=13)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Implementation	3	9	11	44	1	17	2	11	10	56	8	89	2	15
Sustainment	0	0	5	20	0	0	1	6	2	11	1	11	8	62
Not specified	29	91	9	36	5	83	15	83	6	33	0	0	3	23
Outcomes assessed														
Acceptability	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Appropriateness	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Adoption	2	6	0	0	0	0	0	0	0	0	0	0	0	0
Cost	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Feasibility	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fidelity	3	9	0	0	0	0	1	6	1	6	1	11	0	0
Penetration	1	3	0	0	1	17	0	0	0	0	0	0	0	0
Sustainability	2	6	0	0	0	0	0	0	0	0	1	11	2	15

EPIS: exploration, adoption/preparation, implementation, sustainment.

There was no information available on any of the remaining psychometric criteria. It is worth noting that there was only one scale relevant to feasibility in this measure and the scores above are rolled up to reflect the score for the broader measure.

**Fidelity.** Eighteen measures of *fidelity* were identified in mental or behavioral health research. Information about internal consistency was available for six measures, convergent validity for three measures, discriminant validity for two measures, concurrent validity one measure, predictive validity for four measures, known-groups validity for two measures, structural validity for two measures, responsiveness for no measures, and norms for nine measures. For those measures of *feasibility* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “3—good,” “3—good” for convergent validity, “1—minimal/emerging” for discriminant validity, “1—minimal/emerging” for concurrent validity, “3—good” for predictive validity, “1—minimal/emerging” for known-groups validity, “3—good” for structural validity, and “2—adequate” for norms.

The Yale Adherence and Competence Scale had the highest psychometric rating score among measures of *fidelity* used in mental and behavioral health care (psychometric total maximum score=9; maximum possible score=36), with ratings of “4—excellent” for convergent validity, “1—minimal/emerging” for discriminant validity, “1—minimal/emerging” for concurrent validity, and “3—good” for structural validity (Carroll et al., 2000). There was no information available on any of the remaining psychometric criteria.

**Penetration.** Twenty-three measures of *penetration* were identified in mental or behavioral health research of which

14 were deemed unsuitable for rating. Information about internal consistency was available for three measures, convergent validity for one measure, discriminant validity for no measures, concurrent validity for no measures, predictive validity for two measures, known-groups validity for no measures, structural validity for one measure, responsiveness for one measure, and norms for three measures. For those measures of *penetration* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “3—good,” “3—good” for convergent validity, “2—adequate” for predictive validity, “1—minimal/emerging” for structural validity, and “1—minimal/emerging” for norms.

The Degree of Implementation Form had the highest psychometric rating score among measures of *penetration* used in mental and behavioral health care (psychometric total maximum score=6; maximum possible score=36), with ratings of “2—adequate,” for internal consistency and “4—excellent” for convergent validity (Forchuk et al., 2002). There was no information available on any of the remaining psychometric criteria.

**Sustainability.** Fourteen measures of *sustainability* were identified in mental or behavioral health research of which one was deemed unsuitable for rating. Information about internal consistency was available for six measures, convergent validity for one measure, discriminant validity for no measures, concurrent validity for three measures, predictive validity for two measures, known-groups validity for three measures, structural validity for four measures, responsiveness for no measures, and norms for eight measures. For those measures of *penetration* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,” “4—excellent” for convergent

**Table 4.** Data availability.

Construct	N	Internal consistency		Convergent validity		Discriminant validity		Known-groups validity		Predictive validity		Concurrent validity		Structural validity		Responsiveness		Norms	
		n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Acceptability	32	19	59	10	31	-	-	4	13	6	19	1	3	4	13	2	6	20	63
Adoption	25	16	64	4	16	-	-	2	8	5	20	2	8	4	16	1	4	12	48
Appropriateness	6	3	50	1	17	-	-	-	-	1	17	-	-	1	17	-	-	2	33
Feasibility	18	5	28	-	-	-	-	1	6	1	6	-	-	1	6	-	-	9	50
Fidelity	18	6	33	3	17	2	11	2	11	4	22	1	6	2	11	-	-	9	50
Penetration	9	3	33	1	11	-	-	-	-	2	22	-	-	1	11	1	-	3	33
Sustainability	13	6	46	1	8	-	-	3	23	2	15	3	23	4	30	-	-	8	62

Excluding measures that were deemed unsuitable for rating.

**Table 5.** Psychometric scores summary data.

Construct	N	Internal consistency		Convergent validity		Discriminant validity		Known-groups validity		Predictive validity		Concurrent validity		Structural validity		Responsiveness		Norms	
		M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R
Acceptability	32	3	1-4	2	1-4	-	-	-1	-1-3	-1	-1-4	2	2	-1	-1-4	1	4	1	-1-4
Adoption	25	2	1-4	2	1-3	-	-	-1	-1-1	1	-1-3	2	1-2	1	-1-2	1	1	2	-1-4
Appropriateness	6	4	3-4	3	3	-	-	-	-	3	3	-	-	4	4	-	-	1	-1-3
Feasibility	18	4	2-4	-	-	-	-	-1	-1	1	1	-	-	2	2	-	-	-1	-1-2
Fidelity	18	3	1-4	3	2-4	1	-1-2	1	1-2	3	-1-4	1	1-2	3	1-4	-	-	2	-1-2
Penetration	9	3	2-4	3	3-4	-	-	-	-	2	1-2	-	-	1	1	-	-	1	-1-2
Sustainability	13	2	1-4	4	4	-	-	3	1-3	-1	-1-2	1	1-3	2	1-3	-	-	1	-1-3

Note. M = median; R = range. Excluding zeros where psychometric information not available; excluding measures that were deemed unsuitable for rating; scores range from "poor" (-1) to "excellent" (4).

**Table 6.** Individual psychometric ratings.

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
<b>Acceptability</b>										
Abbreviated Acceptability Rating Profile (Tarnowski & Simonian, 1992)	4	0	0	0	0	0	0	0	1	5
Acceptability of Intervention Measure–posttest (Henninger, 2010)	4	0	0	0	0	0	0	0	2	6
Acceptability of Intervention Measure–pretest (Henninger, 2010)	4	0	0	0	0	0	0	0	2	6
Behavior Intervention Rating Scale (S. N. Elliott & Treuting, 1991)	4	2	0	0	0	0	–1	0	1	6
Behavior Intervention Rating Scale (Acceptability Scale) (S. N. Elliott & Treuting, 1991)	4	2	0	0	0	0	–1	0	–1	4
Behavior Intervention Rating Scale (Effectiveness Scale) (S. N. Elliott & Treuting, 1991)	4	2	0	0	0	0	–1	0	–1	4
Behavior Intervention Rating Scale (Time Scale) (S. N. Elliott & Treuting, 1991)	3	1	0	0	0	0	0	0	0	4
Behavioral Interventionist Satisfaction Survey (K. A. McLean, 2013)	3	0	0	0	–1	0	0	0	–1	1
Behavioral Interventionist Satisfaction Survey (Acceptability/Feasibility) (K. A. McLean, 2013)	2	0	0	0	1	0	0	0	–1	2
Children's Intervention Rating Profile (S. N. Elliott et al., 1986)	3	0	0	0	0	0	0	0	0	3
Children's Usage Rating Profile (Briesch & Chafouleas, 2009)	3	4	0	0	0	0	3	0	2	12
Children's Usage Rating Profile (Personal Desirability Subscale) (Briesch & Chafouleas, 2009)	4	4	0	0	0	0	0	0	2	10
Classroom Organization and Management Program Treatment Acceptability (Tanol, 2010)	4	0	0	0	2	–1	0	0	–1	4
Eisen Provider Knowledge & Attitudes Survey (Eisen et al., 2013)	3	0	0	0	0	3	0	0	–1	5
Evidence-Based Practice Attitudes Scale (Aarons, 2004)	3	2	0	2	1	–1	2	1	2	12
Evidence-Based Practice Attitudes Scale (Appeal Scale) (Aarons, 2004)	2	1	0	0	1	–1	0	1	2	6
Evidence-Based Practice Attitudes Scale (Divergent Scale) (Aarons, 2004)	1	1	0	0	–1	–1	0	0	2	2
Evidence-Based Practice Attitudes Scale (Openness Scale) (Aarons, 2004)	3	2	0	0	1	–1	0	0	2	7
Evidence-Based Practice Attitudes Scale (Requirement Scale) (Aarons, 2004)	4	1	0	0	–1	1	0	0	2	7
Good Behavior Game–response Cost/GBG–reinforcement Treatment Acceptability (Tanol et al., 2010)	0	0	0	0	0	0	0	0	0	0
Influences on Psychotherapy Training Participation Scale (A. Lyon, 2010)	0	0	0	0	0	0	0	0	0	0
Intervention Profile Rating Scale (Kutsick et al., 1991)	1	0	0	0	0	0	0	0	0	0
Intervention Rating Profile-10 (Power et al., 1995)	4	3	0	0	0	0	0	0	2	3
Manual Acceptability Rating Scale (D. Mline, 2010)	0	0	0	0	0	0	0	0	0	0
Modified Treatment Evaluation Inventory (Johnston & Fine, 1993)	3	3	0	0	0	0	0	0	–1	5
Parent Evaluation Inventory (Kazdin et al., 1989)	0	4	0	0	0	0	0	4	–1	7
Parent Evaluation Inventory (Acceptability) (Kazdin et al., 1989)	0	4	0	0	0	0	0	0	0	4
Parent Evaluation Inventory (Parent Progress) (Kazdin et al., 1989)	0	4	0	0	0	0	0	0	0	4
Parenting Strategies Questionnaire (Whittingham et al., 2006)	4	0	0	0	3	0	0	0	–1	6
Parenting Strategies Questionnaire (Acceptability Scale) (Whittingham et al., 2006)	4	0	0	0	0	0	0	0	–1	3
Perceptions of Computerized Therapy Questionnaire—Clinician Version (Carper et al., 2013)	0	0	0	0	4	0	0	0	–1	3
SFTRC Course Evaluation (Attitudes—Negative Outcomes) (Carper et al., 2013)	0	0	0	0	0	0	0	0	2	2
SFTRC Course Evaluation (Attitudes—Positive Outcomes) (Carper et al., 2013)	0	0	0	0	0	0	0	0	2	2
Suicide Prevention Program Rating Profile (Eckert et al., 2003)	4	0	0	0	0	0	0	0	4	8
The LOCI Feasibility, Acceptability, Utility Scale (Aarons et al., 2015)	0	0	0	0	0	0	0	0	–1	–1
The LOCI Feasibility, Acceptability, Utility Scale (Acceptability) (Aarons et al., 2015)	0	0	0	0	0	0	0	0	–1	–1

(Continued)

**Table 6. (Continued)**

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
The Practitioners Attitudes Toward Treatment Manuals (Addis & Krasnow, 2000)	3	3	0	0	1	0	0	0	4	11
The Practitioners Attitudes Toward Treatment Manuals (Negative Process) (Addis & Krasnow, 2000)	3	3	0	0	0	0	0	0	4	10
The Practitioners Attitudes Toward Treatment Manuals (Positive Outcome) (Addis & Krasnow, 2000)	4	3	0	0	0	0	0	0	4	11
The Pre-referral Intervention Team Inventory (Yetter, 2010)	4	3	0	0	0	0	4	0	3	14
The Pre-referral Intervention Team Inventory (Acceptability Subscale) (Yetter, 2010)	4	3	0	0	0	0	0	0	0	7
The Pre-referral Intervention Team Inventory (Effectiveness Subscale) (Yetter, 2010)	4	3	0	0	0	0	0	0	0	7
Therapist Evaluation Inventory—Parent Management Training (Myers, 2008)	0	4	0	0	0	0	0	0	-1	3
Therapist Evaluation Inventory—Parent Management Training (Parent Improvement) (Myers, 2008)	0	4	0	0	0	0	0	0	0	4
Therapist Evaluation Inventory—Parent Management Training (Parent Progress) (Myers, 2008)	0	4	0	0	0	0	0	0	0	4
Training Acceptability Rating Scale (D. Milne, 2010)	0	0	0	0	0	-1	0	0	1	0
Training Acceptability Rating Scale (Feasibility and Acceptability Scale) (D. Milne, 2010)	0	0	0	0	0	-1	0	0	1	0
Training Acceptability Rating Scale (Satisfaction Scale) (D. Milne, 2010)	0	0	0	0	0	-1	0	0	1	0
Training Acceptability Rating Scale-1 (Davis & Capponi, 1989)	3	4	0	0	0	0	0	0	1	8
Training Acceptability Rating Scale-2 (D. Milne & Noone, 1996)	0	0	0	0	0	0	0	0	0	0
Training/Practice Acceptability/Feasibility/Appropriateness Scale (A. Lyon, 2011)	0	0	0	0	0	0	0	0	0	0
Treatment Acceptability Rating Form (Reimers, 1988)	0	0	0	0	0	0	0	0	0	0
Treatment Evaluation Inventory—Short Form (Newton & Sturmeay, 2004)	4	0	0	0	0	0	0	0	-1	3
Treatment Plan Implementation Monitoring and Feedback Acceptability Survey (Easton, 2009)	0	0	0	0	0	0	0	0	0	0
Adoption										
Adoption of Smoking Cessation Expert System (Hoving et al., 2006)	0	0	0	0	0	0	0	0	0	0
Adoption of the Principles of Effectiveness Survey (Pankratz et al., 2002)	3	0	0	0	0	0	0	0	0	3
ASE Determinants Questionnaire (Zwerver et al., 2013)	2	0	0	0	0	1	0	0	-1	2
ASE Determinants Questionnaire (Intentions Scale) (Zwerver et al., 2013)	2	0	0	0	0	1	0	0	-1	2
Barwick Measure of Adoption of Research (Barwick et al., 2008)	0	0	0	0	0	0	0	0	0	0
BB1 Intent Measure (Brothers et al., 2015)	3	2	0	0	0	0	0	0	2	7
Buckley Stages of Change Tool (Buckley et al., 2003)	0	0	0	0	0	0	0	0	0	0
Cook Measure of Adoption (Cook et al., 2009)	0	0	0	0	0	0	0	0	0	0
David CPP Adoption Measure (David & Schiff, 2015)	0	0	0	0	0	0	0	0	0	0
Evidence-Based Practice Attitudes Scale (Aarons, 2004)	3	2	0	2	1	-1	2	1	2	12
Evidence-Based Practice Attitudes Scale (Appeal Scale) (Aarons, 2004)	2	1	0	0	1	-1	0	1	2	6
Evidence-Based Practice Attitudes Scale (Divergent Scale) (Aarons, 2004)	1	1	0	0	-1	-1	0	0	2	2
Evidence-Based Practice Attitudes Scale (Openness Scale) (Aarons, 2004)	3	2	0	0	1	-1	0	0	2	7
Evidence-Based Practice Attitudes Scale (Requirement Scale) (Aarons, 2004)	4	1	0	0	-1	1	0	0	2	7
Himmelhoch Measure of Tobacco Cessation Intervention Utilization (Himmelhoch et al., 2014)	0	0	0	0	0	0	0	0	0	0
Kelly Intention to Adopt EBP Measure (Kelly & Deane, 2012)	3	0	0	0	0	0	0	0	2	5
Kelly Intention to Adopt EBP Measure (Behavioral) (Kelly & Deane, 2012)	3	0	0	0	0	0	0	0	0	3
Kelly Intention to Adopt EBP Measure (Cognitive) (Kelly & Deane, 2012)	3	0	0	0	0	0	0	0	2	5

(Continued)

**Table 6. (Continued)**

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Kelly Intention to Adopt EBP Measure (Psychodynamic) (Kelly & Deane, 2012)	4	0	0	0	0	0	0	0	0	4
Knowledge Exchange Outcomes Tool (Skinner, 2007)	0	0	0	0	0	0	0	0	0	0
Knowledge Utilization Questionnaire (Chagnon et al., 2010)	0	0	0	0	0	0	0	0	0	0
Moore & Benbasat Adoption of Information Technology Innovation Measure (Moore & Benbasat, 1991)	3	0	0	0	0	0	0	0	0	3
Parenting Strategies Questionnaire (Behavioral Intention Scale) (Moore & Benbasat, 1991)	4	0	0	0	0	0	0	0	-1	3
Perceptions of Computerized Therapy Questionnaire—Clinician Version (Future Use Intentions) (Moore & Benbasat, 1991)	0	0	0	0	0	0	0	0	-1	-1
Prevention Program Assessment (Stamatakis et al., 2012)	2	0	0	0	0	0	1	0	2	5
Prevention Program Assessment (Adoption Scale) (Stamatakis et al., 2012)	3	0	0	0	0	0	0	0	2	5
Prevention Program Assessment (Implementation Scale) (Stamatakis et al., 2012)	2	0	0	0	0	0	0	0	2	4
Ruzeck Measure (Ruzek et al., 2016)	2	0	0	0	3	0	0	0	4	9
Ruzeck Measure (Intention to Use Scale) (Ruzek et al., 2016)	1	0	0	0	0	0	0	0	0	1
School-wide Universal Behavior Sustainability Index—School Teams (District Priority) (Ruzek et al., 2016)	0	0	0	0	1	0	0	0	0	1
School-wide Universal Behavior Sustainability Index—School Teams (School Priority) (Ruzek et al., 2016)	0	0	0	2	-1	0	0	0	0	1
School-wide Universal Behavior Sustainability Index—School Teams (Team Use of Data) (Ruzek et al., 2016)	0	0	0	1	2	0	0	0	0	3
SFTRC Course Evaluation (Haug et al., 2008)	2	0	0	0	0	0	0	0	2	4
SFTRC Course Evaluation (Organizational Barriers) (Haug et al., 2008)	2	0	0	0	0	0	0	0	2	4
TCU Workshop Evaluation Form (Simpson, 2002)	2	0	0	0	1	0	-1	0	2	4
The LOCI Feasibility, Acceptability, Utility Scale (Aarons et al., 2015)	0	0	0	0	0	0	0	0	-1	-1
Walitzer Measure of Adoption Stages (Walitzer et al., 2015)	3	0	0	0	2	0	2	0	0	7
Walitzer Measure of Adoption Stages (Confirmation) (Walitzer et al., 2015)	3	0	0	0	0	0	0	0	0	3
Walitzer Measure of Adoption Stages (Decision) (Walitzer et al., 2015)	3	0	0	0	0	0	0	0	0	3
Walitzer Measure of Adoption Stages (Implementation) (Walitzer et al., 2015)	2	0	0	0	0	0	0	0	0	2
Walitzer Measure of Adoption Stages (Knowledge) (Walitzer et al., 2015)	3	0	0	0	0	0	0	0	0	3
Walitzer Measure of Adoption Stages (Persuasion) (Walitzer et al., 2015)	3	0	0	0	2	0	0	0	0	5
Williams Intention to Adopt Measure (Williams et al., 2017)	3	3	0	0	0	0	0	0	3	9
Workshop Assessment Follow-Up Form (Simpson, 2002)	3	3	0	0	0	0	0	0	0	6
Workshop Assessment Follow-Up Form (Trial Use) (Simpson, 2002)	3	3	0	0	0	0	0	0	0	6
Appropriateness										
The Pre-referral Intervention Team Inventory (Yetter, 2010)	4	3	0	0	0	0	4	0	3	14
Influences on Psychotherapy Training Participation Scale (A. Lyon, 2010)	0	0	0	0	0	0	0	0	0	0
Moore & Benbasat Adoption of Information Technology Innovation Measure (Compatibility Scale) (A. Lyon, 2010)	3	0	0	0	0	0	0	0	0	3
Knowledge Exchange Outcomes Tool (Skinner, 2007)	0	0	0	0	0	0	0	0	0	0
Guerrero "Relevancy" Measure (Guerrero et al., 2016)	0	0	0	0	0	0	0	0	0	0

(Continued)

**Table 6. (Continued)**

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Parenting Strategies Questionnaire (Whittingham et al., 2006)	4	0	0	0	3	0	0	0	-1	6
Feasibility										
Children's Usage Rating Profile (Feasibility Scale) (Whittingham et al., 2006)	4	0	0	0	0	0	0	0	2	6
Training Acceptability Rating Scale (Feasibility and Acceptability Scale) (Whittingham et al., 2006)	0	0	0	0	0	-1	0	0	1	0
Behavioral Interventionist Satisfaction Survey (Acceptability/Feasibility) (Whittingham et al., 2006)	2	0	0	0	1	0	0	0	-1	2
Parenting Strategies Questionnaire (Usability Scale) (Whittingham et al., 2006)	4	0	0	0	0	0	0	0	-1	3
Evital Feasibility Questionnaire (Alonso et al., 2010)	0	0	0	0	0	0	0	0	0	0
Feasibility of Administering the Screening Tool for Autism in Toddlers and Young Children (Kobak et al., 2011)	0	0	0	0	0	0	0	0	-1	-1
Feasibility of the Stages of Recovery Instrument (Weeks et al., 2011)	0	0	0	0	0	0	0	0	0	0
Feasibility Questionnaire for Threshold Assessment Grid (Slade et al., 2001)	0	0	0	0	0	0	0	0	-1	-1
The Measure of Disseminability (Trent, 2010)	3	0	0	0	0	0	2	0	0	5
Feasibility of a Screening Tool (Hides et al., 2007)	0	0	0	0	0	0	0	0	2	2
Questionnaire for User Interface Satisfaction (Chin & Norman, 1988)	4	0	0	0	0	0	0	0	0	4
Ebert Providers' Perception of IC Questionnaire (Ebert et al., 2014)	0	0	0	0	0	0	0	0	0	0
Structured Assessment of Feasibility (Bird et al., 2014)	0	0	0	0	0	0	0	0	0	0
System Usability Scale (Kobak et al., 2015)	0	0	0	0	0	0	0	0	0	0
Readiness Ruler (Herie Version) (Miller & Rollnick, 2002)	0	0	0	0	0	0	0	0	-1	-1
Malte Post-treatment Smoking Cessation Beliefs Measure (Feasibility Scale) (Miller & Rollnick, 2002)	0	0	0	0	0	0	0	0	0	0
Malte Post-treatment Smoking Cessation Beliefs Measure (Malte et al., 2013)	0	0	0	0	0	0	0	0	0	0
The LOCI Feasibility, Acceptability, Utility Scale (Feasibility) (Aarons et al., 2015)	0	0	0	0	0	0	0	0	-1	-1
The LOCI Feasibility, Acceptability, Utility Scale (Aarons et al., 2015)	0	0	0	0	0	0	0	0	-1	-1
Fidelity										
Therapy Procedures Checklist (Psychodynamic) (Weersing et al., 2002)	4	0	0	0	0	0	0	0	0	4
Therapy Procedures Checklist (Cognitive) (Weersing et al., 2002)	3	0	0	0	0	0	0	0	2	5
Therapy Procedures Checklist (Behavioral) (Weersing et al., 2002)	3	0	0	0	0	0	0	0	0	3
Therapy Procedures Checklist (Weersing et al., 2002)	3	0	0	0	0	0	0	0	2	5
Therapy Process Observational Coding Systems for Child Psychotherapy-Revised (Cognitive) (McLeod et al., 2015)	0	4	0	0	0	0	0	0	0	4
Therapy Process Observational Coding Systems for Child Psychotherapy-Revised (Behavioral) (McLeod et al., 2015)	0	4	0	0	0	0	0	0	0	4
Therapy Process Observational Coding Systems for Child Psychotherapy-Revised (Psychodynamic) (McLeod et al., 2015)	0	3	0	0	0	0	0	0	0	3
Therapy Process Observational Coding Systems for Child Psychotherapy-Revised (Family) (McLeod et al., 2015)	0	3	0	0	0	0	0	0	0	3
Therapy Process Observational Coding Systems for Child Psychotherapy-Revised (Client Centered) (McLeod et al., 2015)	0	3	0	0	0	0	0	0	0	3

(Continued)

**Table 6.** (Continued)

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Therapy Process Observational Coding Systems for Child Psychotherapy—Revised (McLeod et al., 2015)	0	3	0	0	0	0	0	0	1	4
Ehrhart Fidelity Measure (Ehrhart et al., 2015)	0	0	0	0	0	0	0	0	0	0
Neumiller Fidelity Measure (Neumiller et al., 2009)	0	0	0	0	0	0	0	0	0	0
Chodosh Competency Survey (Chodosh et al., 2015)	0	0	0	0	0	0	0	0	0	0
ARC Principles Questionnaire (Glisson et al., 2013)	2	0	0	0	3	1	0	0	0	6
Implementation of Treatment Integrity Procedures Scale (Perepletchikova et al., 2007)	0	0	2	0	0	0	0	0	2	4
St. George Fidelity Measure (Adherence) (George et al., 2016)	0	0	0	0	3	0	3	0	0	6
St. George Fidelity Measure (Competence) (George et al., 2016)	0	0	0	0	4	0	3	0	0	7
St. George Fidelity Measure (George et al., 2016)	0	0	0	0	3	0	3	0	1	7
Oxman Fidelity Measure (Oxman et al., 2006)	0	0	0	0	1	0	0	0	2	3
Wolmann Fidelity Instrument (Wolmann et al., 2008)	0	0	0	0	0	0	0	0	-1	-1
General Psychiatric Management Adherence Scale (Assessment Procedures) (Kolla et al., 2009)	4	0	0	0	0	0	0	0	0	4
General Psychiatric Management Adherence Scale (Establishing Treatment Contract) (Kolla et al., 2009)	2	0	0	0	0	0	0	0	0	2
General Psychiatric Management Adherence Scale (Ongoing Case Management) (Kolla et al., 2009)	4	0	0	0	0	0	0	0	0	4
General Psychiatric Management Adherence Scale (General Principles of Psychotherapy) (Kolla et al., 2009)	4	0	0	0	0	0	0	0	0	4
General Psychiatric Management Adherence Scale (Focus on Feelings) (Kolla et al., 2009)	4	0	0	0	0	0	0	0	0	4
General Psychiatric Management Adherence Scale (Specific Therapeutic Issues) (Kolla et al., 2009)	3	0	0	0	0	0	0	0	0	3
General Psychiatric Management Adherence Scale (Kolla et al., 2009)	4	0	0	0	0	0	0	0	0	4
Integrated Treatment Fidelity scale (McKee et al., 2013)	0	0	0	0	0	0	0	0	2	2
Yale Adherence and Competence Scale (Assessment) (Carroll et al., 2000)	0	3	0	1	0	0	4	0	0	8
Yale Adherence and Competence Scale (General Support) (Carroll et al., 2000)	0	4	0	2	0	0	4	0	0	10
Yale Adherence and Competence Scale (Goals of Treatment) (Carroll et al., 2000)	0	4	0	1	0	0	4	0	0	9
Yale Adherence and Competence Scale (CM) (Carroll et al., 2000)	0	0	2	2	0	0	2	0	0	6
Yale Adherence and Competence Scale (TSF) (Carroll et al., 2000)	0	0	1	1	0	0	2	0	0	4
Yale Adherence and Competence Scale (CBT) (Carroll et al., 2000)	0	0	-1	1	0	0	1	0	0	1
Yale Adherence and Competence Scale (Carroll et al., 2000)	0	4	1	1	0	0	3	0	0	9
Primary Care Behavior Health Provider Adherence Questionnaire (Beehler et al., 2013)	4	0	0	0	0	2	0	0	-1	5
Azocar Measure of Guideline Adherence (Azocar et al., 2003)	1	0	0	0	0	0	0	0	0	1
Gifford Fidelity Measure (M) (Gifford et al., 2012)	3	0	0	0	0	0	0	0	2	5
Gifford Fidelity Measure (CBT) (Gifford et al., 2012)	4	0	0	0	0	0	0	0	2	6
Gifford Fidelity Measure (Community Reinforcement) (Gifford et al., 2012)	3	0	0	0	0	0	0	0	2	5
Gifford Fidelity Measure (TSF) (Gifford et al., 2012)	3	0	0	0	0	0	0	0	2	5
Gifford Fidelity Measure (CM) (Gifford et al., 2012)	2	0	0	0	0	0	0	0	2	4
Gifford Fidelity Measure (Couples Therapy) (Gifford et al., 2012)	2	0	0	0	0	0	0	0	2	4
Gifford Fidelity Measure (Gifford et al., 2012)	3	0	0	0	0	0	0	0	2	5

(Continued)

**Table 6. (Continued)**

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Lloyd Fidelity Measure (Lloyd et al., 2015)	0	0	0	0	0	0	0	0	0	0
Chou Measure of Guidelines Information (Chou et al., 2011)	0	2	0	0	-1	0	0	0	0	1
Penetration										
Level of Institutionalization Scale (Goodman et al., 1993)	3	0	0	0	1	0	1	0	0	5
Reach Reporter (Baiastramanian et al., 2015)	0	0	0	0	0	0	0	0	0	0
Dialectical Behavior Therapy Skills Use Self-Report (Dimeff et al., 2011)	0	0	0	0	0	0	0	1	2	3
AF-CBT Implementation Measure (Teaching Processes) (Kolko et al., 2012)	3	0	0	0	2	0	0	0	0	5
AF-CBT Implementation Measure (Abuse Specific Skills) (Kolko et al., 2012)	3	0	0	0	1	0	0	0	0	4
AF-CBT Implementation Measure (General Psychological Skills) (Kolko et al., 2012)	4	0	0	0	2	0	0	0	0	6
AF-CBT Implementation Measure (Kolko et al., 2012)	3	0	0	0	2	0	0	0	0	5
The Service Activity Scale (Tobin et al., 2001)	0	0	0	0	0	0	0	0	0	0
Swain Penetration Measure (Swain et al., 2010)	0	0	0	0	0	0	0	0	-1	-1
Degree of Implementation Form (Peer Support) (Forchuk et al., 2002)	0	3	0	0	0	0	0	0	0	3
Degree of Implementation Form (Continued Support) (Forchuk et al., 2002)	0	3	0	0	0	0	0	0	0	3
Degree of Implementation Form (Forchuk et al., 2002)	2	4	0	0	0	0	0	0	0	6
Level of Integration Measure (Clinical Practice) (Beehler et al., 2015)	0	0	0	0	0	0	0	0	1	1
Level of Integration Measure (Systems Integration) (Beehler et al., 2015)	0	0	0	0	0	0	0	0	1	1
Level of Integration Measure (Beehler et al., 2015)	0	0	0	0	0	0	0	0	1	1
Cohen Measure of Reach (Cohen et al., 2015)	0	0	0	0	0	0	0	0	0	0
Sustainability										
Knowledge Exchange Outcomes Tool (Skinner, 2007)	0	0	0	0	0	0	0	0	0	0
Prevention Program Assessment (Evaluation Maintenance) (Stamatakis et al., 2012)	2	0	0	0	0	0	0	0	2	4
Prevention Program Assessment (Resource Maintenance) (Stamatakis et al., 2012)	1	0	0	0	0	0	0	0	2	3
Prevention Program Assessment (Stamatakis et al., 2012)	2	0	0	0	0	0	1	0	2	5
Sustainability and Spread Activities Questionnaire (Chung et al., 2013)	0	0	0	0	0	0	0	0	0	0
General Organizational Index (Bond et al., 2009)	2	4	0	2	-1	0	0	0	-1	6
School-wide Universal Behavior Sustainability Index-School Teams (McIntosh et al., 2011)	4	0	0	3	2	1	3	0	-1	12
School-wide Universal Behavior Sustainability Index-School Teams (Potential Barriers) (McIntosh et al., 2011)	0	0	0	1	0	0	0	0	0	1
Sustainability Model & Guide (AKA Sustainability Index) (Maher et al., 2007)	0	0	0	0	0	0	0	0	3	3
Evidence Based Practice Sustaining Telephone Survey (Swain et al., 2010)	0	0	0	0	0	0	0	0	1	1
Amodeo Counselor Maintenance Measure (Amodeo et al., 2010)	0	0	0	0	0	0	0	0	0	0
Sustainment Cost Survey (Roundfield and Lang, 2017)	0	0	0	0	0	0	0	0	0	0
Program Sustainability Index (Mancini & Marek, 2004)	2	0	0	1	0	0	2	0	2	7
Program Sustainability Assessment Tool (Luke et al., 2014)	3	0	0	0	0	3	2	0	0	8
Change Process Capability Questionnaire (Solberg et al., 2008)	4	0	0	0	0	3	0	0	1	8
Eisen Provider Knowledge & Attitudes Survey (Sustainability)	0	0	0	0	0	0	0	0	-1	-1

Leadership and Organizational Change for Implementation (LOC); San Francisco Treatment Research Center (SFTRC); Attitude, Social norm, Self-efficacy (ASE); Child Parent Psychotherapy (CPP); Bio-behavioral Intervention (BBi); Texas Christian University (TCU); Cognitive Behavioral Therapy (CBT); Twelve Step Facilitation (TSF); Clinical Management (CM); Motivational Interviewing (MI); Alternatives for Families - A Cognitive Behavioral Therapy (AF-CBT); Also known as (AKA)



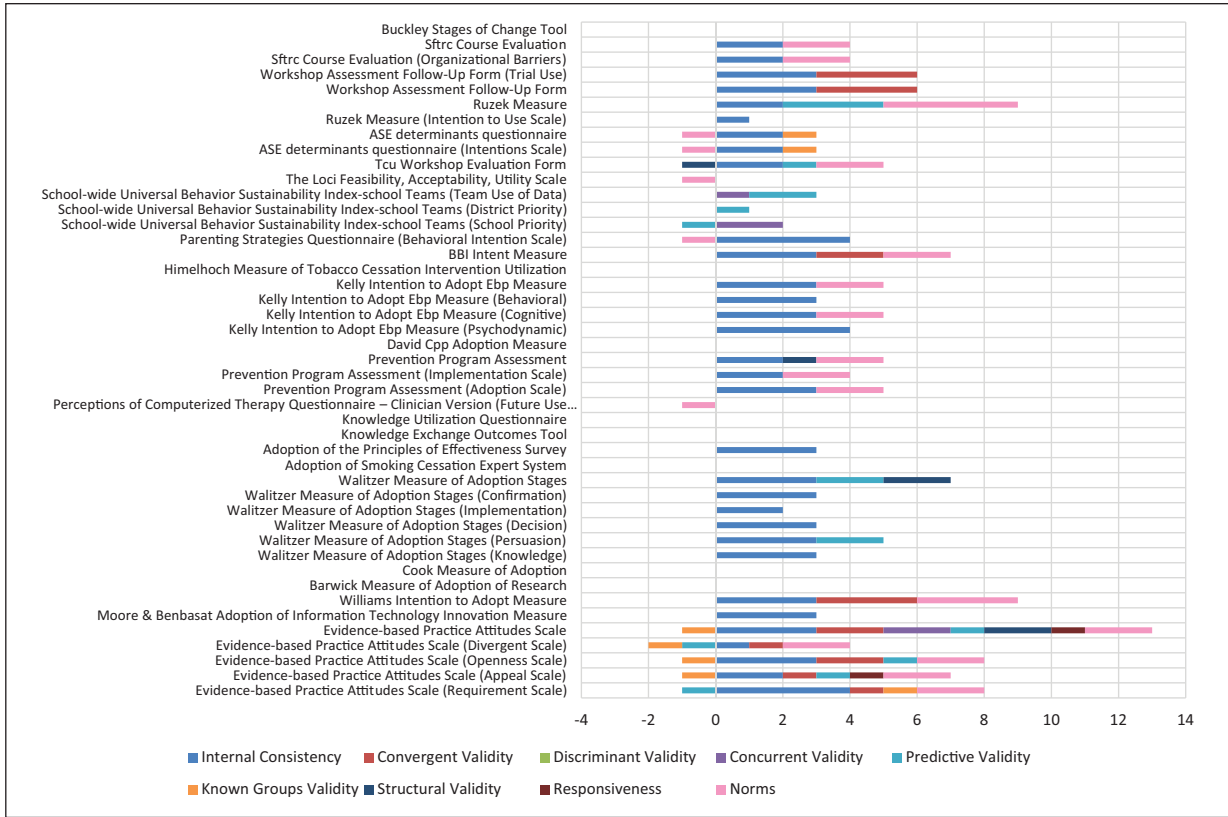


Figure 1. Acceptability ratings.

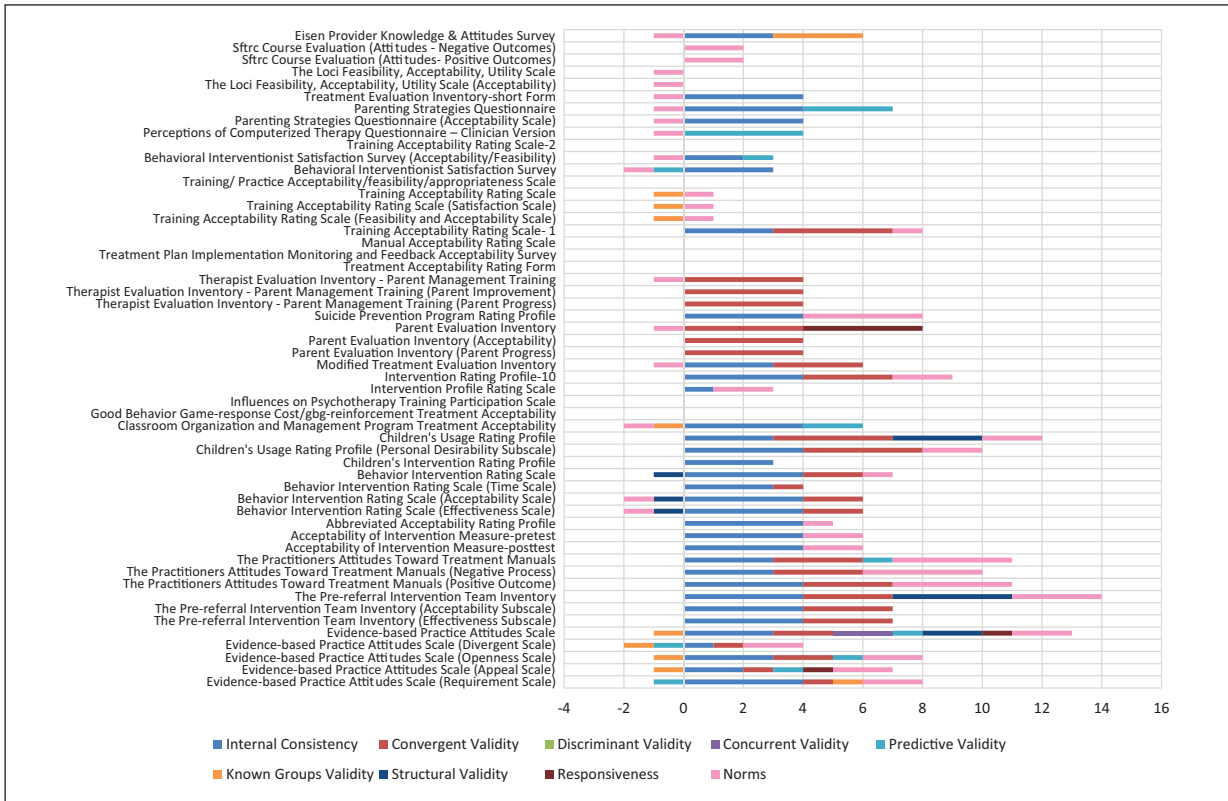


Figure 2. Adoption ratings.

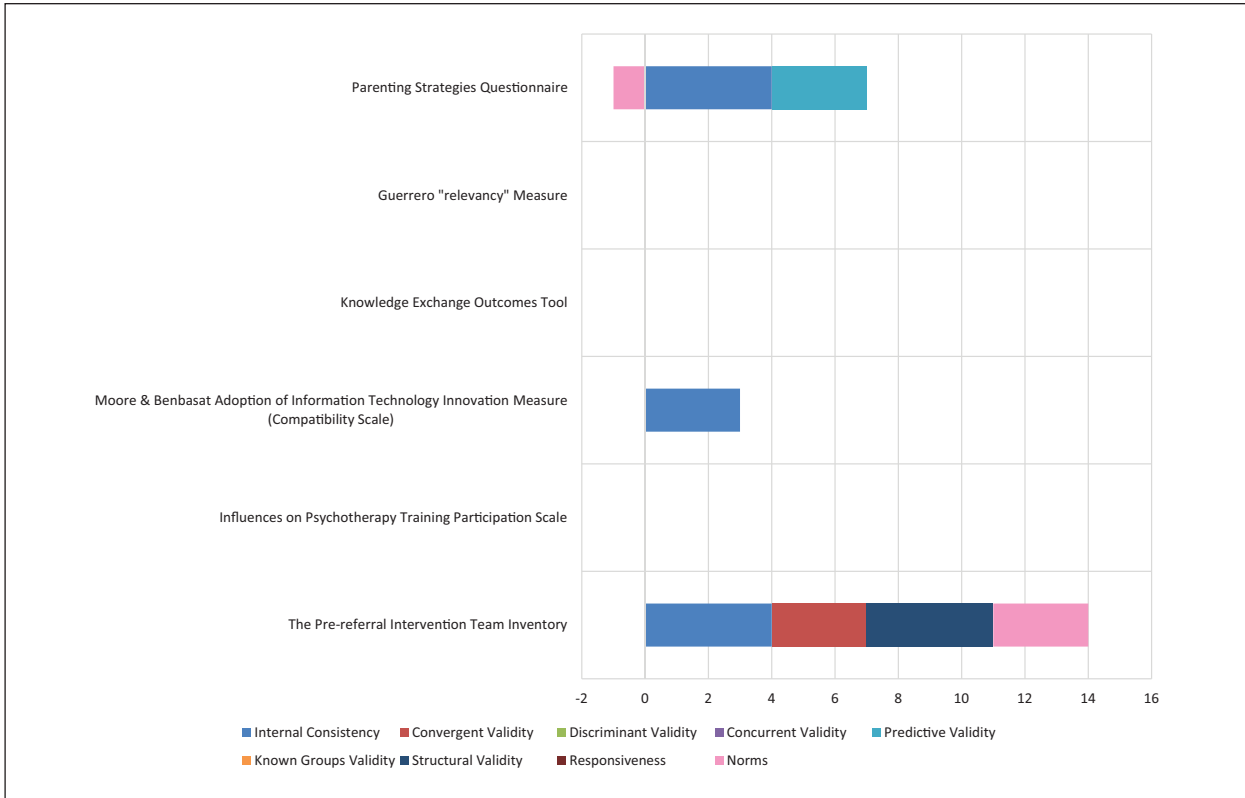


Figure 3. Appropriateness ratings.



Figure 4. Feasibility ratings.

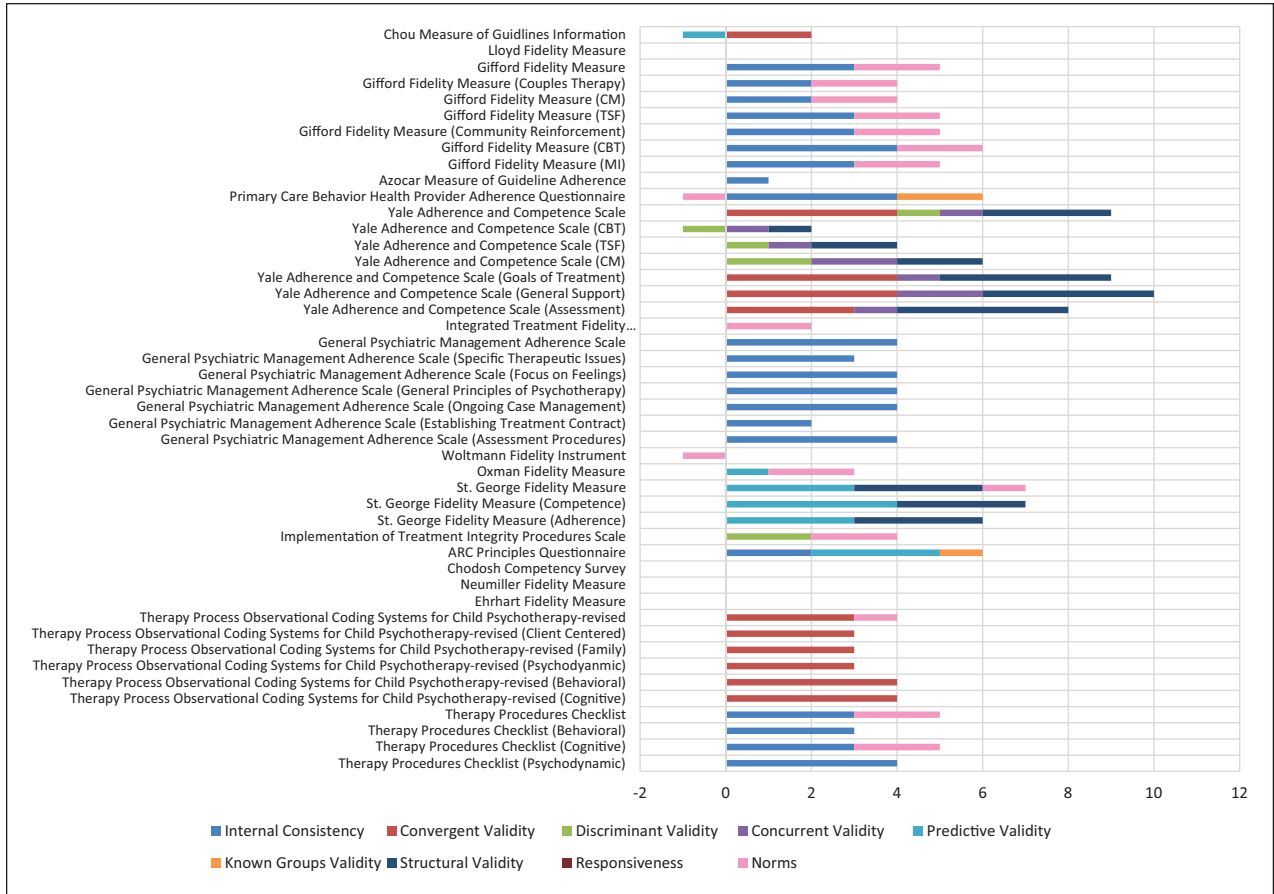


Figure 5. Fidelity ratings.

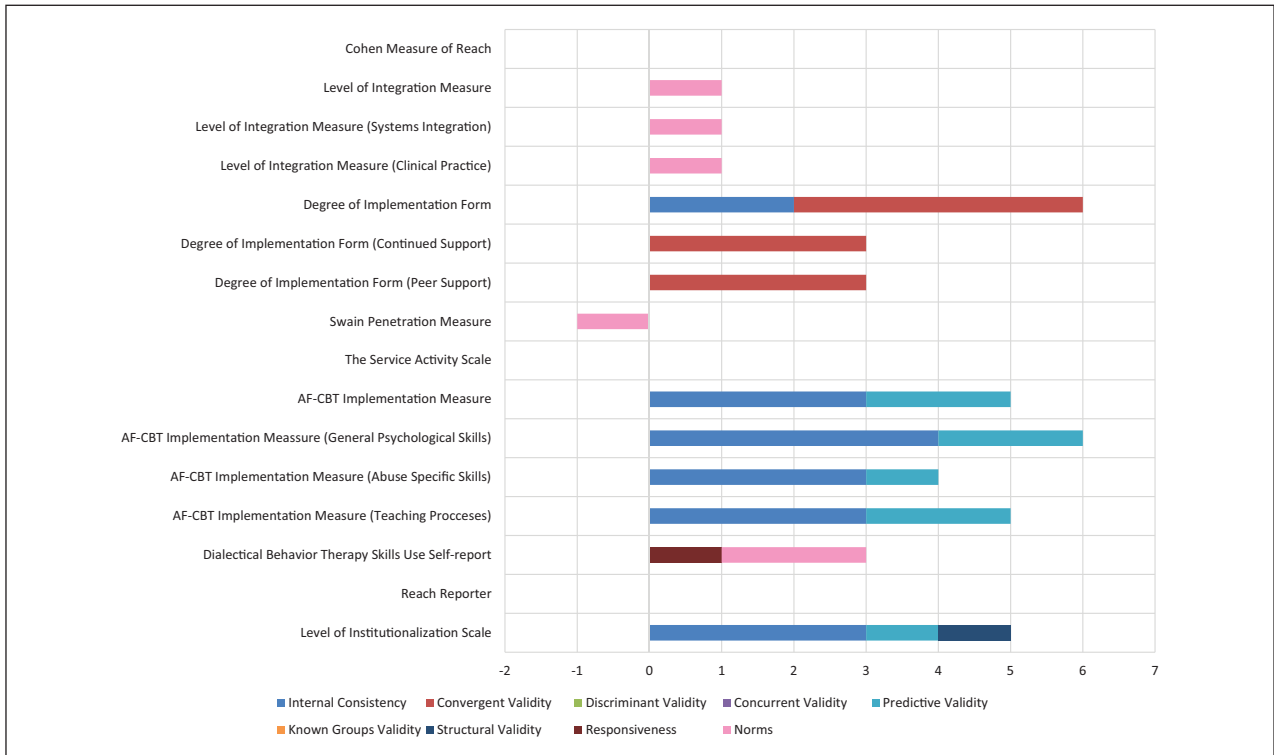


Figure 6. Penetration ratings.



**Figure 7.** Sustainability ratings.

validity, “1—minimal/emerging” for concurrent validity, “-1—poor” for predictive validity, “3—good” for known-groups validity, “2—adequate” for structural validity, and “1—minimal/emerging” for norms.

The School-wide Universal Behavior Sustainability Index-school Teams had the highest psychometric rating score among measures of *penetration* used in mental and behavioral health care (psychometric total maximum score = 12; maximum possible score = 36), with ratings of “4—excellent,” for internal consistency and “3—good” for concurrent validity, “2—adequate” for predictive validity, “1—minimal/emergent” for known-groups validity, “3—good” for structural validity, and “-1—poor” for norms (McIntosh et al., 2011). There was no information available on any of the remaining psychometric criteria.

## Discussion

### Summary of study findings

This systematic review identified 150 measures of implementation outcomes used in mental and behavioral health which were unevenly distributed across the eight outcomes (especially when suitability for rating was concerned). We found 32 measures of *acceptability*, 26 measures of *adoption* (one was deemed unsuitable for rating), 6 measures of

*appropriateness*, 31 measures of cost (none were deemed suitable for rating) 18 measures of *feasibility*, 18 measures of *fidelity*, 23 measures of *penetration* (14 of which were deemed unsuitable for rating), and 14 measures of *sustainability* (one of which was deemed unsuitable for rating). Overall there was limited psychometric information available for measures of implementation outcomes. Norms was the most commonly reported psychometric criterion ( $N=63$ , 52%), followed by internal consistency ( $N=58$ , 48%). Responsiveness was the least reported psychometric property (3%), despite the fact that, for implementation outcomes, responsiveness (or sensitivity to change) is a critically important property. Finally, we found limited evidence of measures’ reliability and validity. Psychometric ratings using the Psychometric and Pragmatic Evaluation Rating Scale (PAPERS; Lewis et al., 2018) ranged from -1 to 14 with a possible minimum of score of -9 and a possible maximum score of 36, illustrating a profound need for measurement studies in this field. This means that almost all measures of implementation outcomes have no evidence (dis)confirming their capacity to detect real change over time, a gap in the literature that requires significant attention and resources.

Measures were moderately generalizable across populations with the majority of empirical uses occurring in studies providing treatment for general mental health

issues, substance use, depression, and other behavioral disorders ( $n=84$ , 63%). The remaining empirical uses occurred in studies evaluating treatments for trauma, anxiety, suicidal ideation, mania, eating disorders, alcohol use, or grief, and several where the population was not specified or did not fit cleanly into any of these categories. Over half of the measure uses occurred in outpatient or school settings ( $n=64$ , 57%), which are the primary settings in which people receive behavioral health care. The remaining empirical uses were in residential, inpatient, state mental health settings, or were not specified.

### *Comparison with previous systematic review*

The findings of this updated review suggest a proliferation of measure development for mental and behavioral health in just the past 2 years—66 new measures were identified—with a continued uneven distribution of measures across implementation outcomes. This demonstrated growth in number of measures confirms that significant focus is being dedicated to measuring implementation outcomes. Importantly, those outcomes that some may argue are relatively unique to implementation, such as feasibility, compared with those that are common to intervention, such as acceptability, increased by 10-fold from 2015 to 2019. It is also worth noting that we found fewer measures of acceptability ( $n=32$ ) than in the previous review ( $n=50$ ). We believe that this discrepancy in identified measures may be due to our refined set of search terms and more structured construct mapping exercise. That is, we used a more precise definition of *acceptability* of an evidence-based practice, as opposed to a looser interpretation of satisfaction with organizational processes, which narrowed our set of synonyms included in the search. In this updated review, construct assignment was checked and confirmed by content expert (CCL) who reviewed items within each measure and/or scale, which was not a process we were able to employ in the initial publication given limits in our funding. This careful item-level may also explain why some measures identified in our initial systematic review were screened out in this updated study.

While more measures in this new review had psychometric information available (86; 70%) on at least one criterion compared with the measures in the previous review (56; 56%), psychometric information for some criteria, such as discriminant, convergent, and concurrent validity as well as responsiveness, remained limited even despite their criticality for scientific evaluations of implementation efforts. We hope that future adoption of measurement reporting standards prompts more reporting and, perhaps, more psychometric testing. However, overall, this finding illustrates that the field is continuing to grow in its testing and reporting of psychometric properties with more attention to the production of valid and reliable measures. With continued focus on gathering information and evidence for

these important psychometric properties, the field may move toward a consensus battery of implementation outcomes measures that can be used across studies to accumulate evidence about what strategies work best for which interventions, for whom, and under what conditions.

The development of “in-house” measures used only once for a specific study contributes to the proliferation of measures that have limited evidence of reliability and validity (Martinez et al., 2014). These measures are typically designed to suit immediate needs of a project and not developed with supportive theory. Of the 150 measures identified, 126 (83%) were only used once in behavioral health care (this included all 48 of the measures deemed not suitable for rating). Of those 78 remaining measures that were suitable for rating, 26 (33%) had available information about internal consistency and scores ranged from “1—minimal/emerging” to “4—excellent.” For convergent validity, seven (9%) measures had information available with scores ranging from “2—adequate” to “4—excellent.” None of these measures had information available for discriminant validity. For concurrent validity, only one (1%) measure had information available and it scored a “1—minimal/emerging.” For predictive validity, 17 (22%) measures had information available with scores ranging from “-1—poor” to “4—excellent.” Only two (3%) measures had information available for known-groups validity and scores ranged from “-1—poor” to “1—minimal/emerging.” For structural validity, only 10 (13%) measures had information and scores ranged from “1—minimal/emerging” to “4—excellent.” For responsiveness, only one (1%) measure had information available and it scored a “1—minimal/emerging.” Finally, 29 (37%) measures had information for norms with scores ranging from “-1—poor” to “4—excellent.” Limiting development of “in-house” measures will also likely increase the ability to accumulate knowledge across studies by deploying common measures.

### *Limitations*

There are several noteworthy limitations to this systematic review. One limitation of the current study is the length of time that has transpired since the original literature searches were completed in 2017. Due to the immense undertaking of the full scope of this R01 project, it took the research team nearly 2 years to screen articles, extract data, apply our rating system, and complete this article. This systematic review is part of a larger project to identify measures of all implementation constructs associated with the Consolidated Framework for Implementation Research Constructs (CFIR) (Damschroder et al., 2009), which are included in this special section. In total, our team conducted 47 systematic reviews over the course of 4 years. Due to this gap in time between when we conducted our searches and when we finalized our data, it is possible that

new measures of implementation outcomes were developed that we did not identify. Despite this, our measure forward “cited-by” searches described above were conducted in the early months of 2019, which gives us confidence that we captured all recent uses of the measures we identified in 2017.

Another limitation is that this review focused only on implementation outcomes in mental and behavioral health care. It is possible that reliable and valid measures of these outcomes exist that have not yet been used in this context. It is also possible that some of the measures included in this review have been used outside of mental health or behavioral health care; in that case, the psychometric ratings described above could change, either positively or negatively, with additional evidence from such studies. A coinciding measurement review (Khadjesari et al., 2017) is underway to identify measures of implementation outcomes in physical health care settings. It will be illuminating to discover how their findings compare with the findings in this review.

Finally, poor reporting practices in published articles not only limited the information available about measures’ psychometric properties, but also the completeness of that information for psychometric rating. For example, authors occasionally reported that structural validity was assessed through exploratory or confirmatory factor analysis; however, they did not report the variance explained by the factors or neglected to report key model fit statistics. Likewise, authors sometimes stated that all scales exhibited internal consistency above a certain threshold (e.g.,  $\alpha > .70$ ) rather than stating the exact values. As is the case in all systematic reviews, the consistency and quality of reporting of measurement properties in the included studies influenced the extent to which measures’ psychometric properties could be rated and the level of uncertainty of those ratings. Relatedly, it is worth noting that although some measurement properties could have improved over time through adaptation and refinement, earlier evidence was also considered in our rating summary which may have negatively skewed the quality of burgeoning measures.

## Conclusion

This systematic measure review highlights significant progress with respect to the development of implementation outcome measures and assessment of their psychometric properties. Even so, our review makes clear the need for additional measure development and testing both to correct the mal-distribution of measures of implementation outcomes and to enhance the psychometric properties of existing measures. Although some of the measures included in this review are promising and merit further refinement and evaluation, most measures lacked information on critical psychometric properties making it unclear whether they warrant further investment. High-quality

measures of outcomes are especially critical for advancing implementation science. A concerted, coordinated effort to develop such measures is needed to gain confidence in the findings of future evaluation efforts.

## Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Funding for this study came from the National Institute of Mental Health, awarded to Dr. Cara Lewis as principal investigator. Dr. Lewis is both an author of this article and an editor of the journal, *Implementation Research and Practice*. Due to this conflict, Dr. Lewis was not involved in the editorial or review process for this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of Mental Health (NIMH) “Advancing implementation science through measure development and evaluation” (1R01MH106510), awarded to Dr. Cara Lewis as principal investigator.

## ORCID iDs

Kayne Mettert  <https://orcid.org/0000-0003-1750-7863>

Cara Lewis  <https://orcid.org/0000-0001-8920-8075>

## References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61–74. <https://doi.org/10.1023/b:mhsr.0000024351.12294.65>
- Aarons, G. A., Ehrhart, M. G., Farahnak, L. R., & Hurlburt, M. S. (2015). Leadership and organizational change for implementation (LOCI): A randomized mixed method pilot study of a leadership and organization development intervention for evidence-based practice implementation. *Implementation Science, 10*(1), Article 11. <https://doi.org/10.1186/s13012-014-0192-y>
- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4–23. <https://doi.org/10.1007/s10488-010-0327-7>
- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists’ attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*(2), 331–339. <https://doi.org/10.1037/0022-006x.68.2.331>
- Alonso, F., Walsh, C. O., Salvador-Carulla, L., & the eVITAL group. (2010). Methodology for the development of a taxonomy and toolkit to evaluate health-related habits and lifestyle (eVITAL). *BMC Research Notes, 3*, Article 83. <https://doi.org/10.1186/1756-0500-3-83>
- Amodeo, M., Storti, S. A., & Larson, M. J. (2010). Moving empirically supported practices to addiction treatment

- programs: Recruiting supervisors to help in technology transfer. *Substance Use & Misuse*, 45(6), 968–982. <https://doi.org/10.3109/10826080903534467>
- Azocar, F., Cuffel, B., Goldman, W., & McCarter, L. (2003). The impact of evidence-based guideline dissemination for the assessment and treatment of major depression in a managed behavioral health care organization. *The Journal of Behavioral Health Services & Research*, 30(1), 109–118. <https://doi.org/10.1007/bf02287816>
- Balasubramanian, B. A., Cohen, D. J., Davis, M. M., Gunn, R., Dickinson, L. M., Miller, W. L., . . . Stange, K. C. (2015). Learning Evaluation: Blending quality improvement and implementation research methods to study healthcare innovations. *Implementation Science*, 10, Article 31. <https://doi.org/10.1186/s13012-015-0219-z>
- Barwick, M. A., Boydell, K. M., Stasiulis, E., Ferguson, H. B., Blase, K., & Fixsen, D. (2008). Research utilization among children’s mental health providers. *Implementation Science*, 3, Article 19. <https://doi.org/10.1186/1748-5908-3-19>
- Beehler, G. P., Funderburk, J. S., King, P. R., Wade, M., & Possemato, K. (2015). Using the Primary Care Behavioral Health Provider Adherence Questionnaire (PPAQ) to identify practice patterns. *Translational Behavioral Medicine*, 5(4), 384–392. <https://doi.org/10.1007/s13142-015-0325-0>
- Beehler, G. P., Funderburk, J. S., Possemato, K., & Vair, C. L. (2013). Developing a measure of provider adherence to improve the implementation of behavioral health services in primary care: A Delphi study. *Implementation Science*, 8, Article 19. <https://doi.org/10.1186/1748-5908-8-19>
- Bird, V. J., Le Boutillier, C., Leamy, M., Williams, J., Bradstreet, S., & Slade, M. (2014). Evaluating the feasibility of complex interventions in mental health services: Standardised measure and reporting guidelines. *The British Journal of Psychiatry*, 204(4), 316–321. <https://doi.org/10.1192/bjp.bp.113.128314>
- Bond, G. R., Drake, R. E., Rapp, C. A., McHugo, G. J., & Xie, H. (2009). 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>
- Briesch, A. M., & Chafouleas, S. M. (2009). Exploring student buy-in: Initial development of an instrument to measure likelihood of children’s intervention usage. *Journal of Educational and Psychological Consultation*, 19(4), 321–336. <https://doi.org/10.1080/10474410903408885>
- Brothers, B. M., Carpenter, K. M., Shelby, R. A., Thornton, L. M., Frierson, G. M., Patterson, K. L., & Andersen, B. L. (2015). Dissemination of an evidence-based treatment for cancer patients: Training is the necessary first step. *Translational Behavioral Medicine*, 5(1), 103–112. <https://doi.org/10.1007/s13142-014-0273-0>
- Buckley, L. L., Goering, P., Parikh, S. V., Butterill, D., & Foo, E. K. (2003). Applying a “stages of change” model to enhance a traditional evaluation of a research transfer course. *Journal of Evaluation in Clinical Practice*, 9(4), 385–390. <https://doi.org/10.1046/j.1365-2753.2003.00407.x>
- Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children*, 63(4), 513–521. <https://doi.org/10.1177/001440299706300406>
- Carper, M. M., McHugh, R. K., & Barlow, D. H. (2013). The dissemination of computer-based psychological treatment: A preliminary analysis of patient and clinician perceptions. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(2), 87–95. <https://doi.org/10.1007/s10488-011-0377-5>
- Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., . . . Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, 57(3), 225–238. [https://doi.org/10.1016/s0376-8716\(99\)00049-6](https://doi.org/10.1016/s0376-8716(99)00049-6)
- Chagnon, F., Pouliot, L., Malo, C., Gervais, M. J., & Pigeon, M. E. (2010). Comparison of determinants of research knowledge utilization by practitioners and administrators in the field of child and family social services. *Implementation Science*, 5, Article 41. <https://doi.org/10.1186/1748-5908-5-41>
- Chin, J. P. D. V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 213–218). Association for Computing Machinery. <https://doi.org/10.1145/57167.57203>
- Chodosh, J., Price, R. M., Cadogan, M. P., Damron-Rodriguez, J., Osterweil, D., Czerwinski, A., . . . Frank, J. C. (2015). A practice improvement education program using a mentored approach to improve nursing facility depression care-preliminary data. *Journal of the American Geriatrics Society*, 63(11), 2395–2399. <https://doi.org/10.1111/jgs.13775>
- Chou, A. F., Vaughn, T. E., McCoy, K. D., & Doebbeling, B. N. (2011). Implementation of evidence-based practices: Applying a goal commitment framework. *Health Care Management Review*, 36(1), 4–17. <https://doi.org/10.1097/HMR.0b013e3181dc8233>
- Chung, H., Duffy, F. F., Katzelnick, D. J., Williams, M. D., Trivedi, M. H., Rae, D. S., & Regier, D. A. (2013). Sustaining practice change one year after completion of the national depression management leadership initiative. *Psychiatric Services*, 64(7), 703–706. <https://doi.org/10.1176/appi.ps.201200227>
- Cohen, D. J., Balasubramanian, B. A., Davis, M., Hall, J., Gunn, R., Stange, K. C., . . . Miller, B. F. (2015). Understanding care integration from the ground up: Five organizing constructs that shape integrated practices. *The Journal of the American Board of Family Medicine*, 28(Suppl 1), S7–S20. <https://doi.org/10.3122/jabfm.2015.S1.150050>
- Cook, J. M., Schnurr, P. P., Biyanova, T., & Coyne, J. C. (2009). Apples don’t fall far from the tree: Influences on psychotherapists’ adoption and sustained use of new therapies. *Psychiatric Services*, 60(5), 671–676. <https://doi.org/10.1176/appi.ps.60.5.671>
- Damschroder, L., Aron, D., Keith, R., Kirsh, S., Alexander, J., & Lowery, J. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, Article 50. <https://doi.org/10.1186/1748-5908-4-50>
- David, P., & Schiff, M. (2015). Learning from bottom-up dissemination: Importing an evidence-based trauma intervention for infants and young children to Israel. *Evaluation and*

- Program Planning*, 53, 18–24. <https://doi.org/10.1016/j.evalprogplan.2015.07.012>
- Davis, J. R. R. E. P., & Capponi, D. R. (1989). Acceptability of behavioral staff management techniques. *Behavioral Interventions*, 4(1), 23–44. <https://doi.org/10.1002/bin.2360040104>
- Dimeff, L. A., Woodcock, E. A., Harned, M. S., & Beadnell, B. (2011). Can dialectical behavior therapy be learned in highly structured learning environments? Results from a randomized controlled dissemination trial. *Behavior Therapy*, 42(2), 263–275. <https://doi.org/10.1016/j.beth.2010.06.004>
- Easton, J. E. (2009). *Teacher acceptability of treatment plan implementation monitoring and feedback methods* [Unpublished master's thesis]. North Carolina State University, Raleigh, NC.
- Ebert, L., Malte, C., Hamlett-Berry, K., Beckham, J., McFall, M., & Saxon, A. (2014). Use of a learning collaborative to support implementation of integrated care for smoking cessation for veterans with posttraumatic stress disorder. *American Journal of Public Health*, 104(10), 1935–1942. <https://doi.org/10.2105/AJPH.2013.301776>
- Eckert, T. L., Miller, D. N., DuPaul, G. J., & Riley-Tillman, T. C. (2003). Adolescent suicide prevention: School psychologists' acceptability of school-based programs. *School Psychology Review*, 32(1), 57–76.
- Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2015). Going above and beyond for implementation: The development and validity testing of the Implementation Citizenship Behavior Scale (ICBS). *Implementation Science*, 10, Article 65. <https://doi.org/10.1186/s13012-015-0255-8>
- Eisen, J. C., Marko-Holguin, M., Fogel, J., Cardenas, A., Bahn, M., Bradford, N., . . . Van Voorhees, B. W. (2013). Pilot Study of Implementation of an Internet-Based Depression Prevention Intervention (CATCH-IT) for adolescents in 12 U.S. primary care practices: Clinical and management/organizational perspectives. *The Primary Care Companion for CNS Disorders*, 15(6), e1–e11. <https://doi.org/10.4088/PCC.10m01065>
- Elliott, S. N., & Treuting, M. V. B. (1991). The Behavior Intervention Rating Scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology*, 29, 43–51. [https://doi.org/10.1016/0022-4405\(91\)90014-I](https://doi.org/10.1016/0022-4405(91)90014-I)
- Elliott, S. N., Witt, J. C., Galvin, G. A., & Moe, G. L. (1986). Children's involvement in intervention selection: Acceptability of interventions for misbehaving peers. *Professional Psychology: Research and Practice*, 17(3), 235–241. <https://doi.org/10.1037/0735-7028.17.3.235>
- Forchuk, C., Hartford, K., Blomqvist, A., Martin, M., Chan, L., & Donner, A. (2002). *Therapeutic relationships: From hospital to community: First year results*. Ottawa Report to the Canadian Health Services Research Foundation. Lawson Health Research Institute, London, ON.
- Fortney, J., Enderle, M., McDougall, S., Clothier, J., Otero, J., Altman, L., & Curran, G. (2012). Implementation outcomes of evidence-based quality improvement for depression in VA community based outpatient clinics. *Implementation Science*, 7, Article 30. <https://doi.org/10.1186/1748-5908-7-30>
- George, S. M. S., Huang, S., Vidot, D. C., Smith, J. D., Brown, C. H., & Prado, G. (2016). Factors associated with the implementation of the Familias Unidas intervention in a type 3 translational trial. *Translational Behavioral Medicine*, 6(1), 105–114. <https://doi.org/10.1007/s13142-015-0344-x>
- Gifford, E. V., Tavakoli, S., Weingardt, K. R., Finney, J. W., Pierson, H. M., Rosen, C. S., & Curran, G. M. (2012). How do components of evidence-based psychological treatment cluster in practice? A survey and cluster analysis. *Journal of Substance Abuse Treatment*, 42(1), 45–55. <https://doi.org/10.1016/j.jsat.2011.07.008>
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, 89(9), 1322–1327. <https://doi.org/10.2105/AJPH.89.9.1322>
- Glisson, C., Hemmelgarn, A., Green, P., & Williams, N. J. (2013). Randomized trial of the Availability, Responsiveness and Continuity (ARC) organizational intervention for improving youth outcomes in community mental health programs. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(5), 493–500. <https://doi.org/10.1016/j.jaac.2013.02.005>
- Goodman, R. M., McLeroy, K. R., Steckler, A. B., & Hoyle, R. H. (1993). Development of level of institutionalization scales for health promotion programs. *Health Education Quarterly*, 20(2), 161–178. <https://doi.org/10.1177/109019819302000208>
- Guerrero, E. G., Padwa, H., Fenwick, K., Harris, L. M., & Aarons, G. A. (2016). Identifying and ranking implicit leadership strategies to promote evidence-based practice implementation in addiction health services. *Implementation Science*, 11, Article 69. <https://doi.org/10.1186/s13012-016-0438-y>
- Haug, N. A., Shopshire, M., Tajima, B., Gruber, V., & Guydish, J. (2008). Adoption of evidence-based practices among substance abuse treatment providers. *Journal of Drug Education*, 38(2), 181–192. <https://doi.org/10.2190/DE.38.2.f>
- Henninger, K. (2010). *Exploring the relationship between factors of implementation, treatment integrity, and reading fluency*. University of Massachusetts Amherst.
- Hides, L., Lubman, D. I., Elkins, K., Catania, L. S., & Rogers, N. (2007). Feasibility and acceptability of a mental health screening tool and training programme in the youth alcohol and other drug (AOD) sector. *Drug and Alcohol Review*, 26(5), 509–515. <https://doi.org/10.1080/09595230701499126>
- Himelhoch, S., Riddle, J., & Goldman, H. H. (2014). Barriers to implementing evidence-based smoking cessation practices in nine community mental health sites. *Psychiatric Services*, 65(1), 75–80. <https://doi.org/10.1176/appi.ps.201200247>
- Hoving, C., Mudde, A. N., & de Vries, H. (2006). Intention to adopt a smoking cessation expert system within a self-selected sample of Dutch general practitioners. *European Journal of Cancer Prevention*, 15(1), 82–86. <https://doi.org/10.1097/01.cej.0000186633.81753.8b>
- Johnston, C., & Fine, S. (1993). Methods of evaluating methylphenidate in children with attention deficit hyperactivity disorder: Acceptability, satisfaction, and compliance. *Journal of Pediatric Psychology*, 18(6), 717–730. <https://doi.org/10.1093/jpepsy/18.6.717>
- Kazdin, A. E., Bass, D., Siegel, T., & Thomas, C. (1989). Cognitive-behavioral therapy and relationship therapy in



- the treatment of children referred for antisocial behavior. *Journal of Consulting and Clinical Psychology*, 57(4), 522–535. <https://doi.org/10.1037//0022-006x.57.4.522>
- Kelly, P. J., & Deane, F. P. L. M. J. (2012). Using the theory of planned behavior to examine residential substance abuse workers intention to use evidence-based practices. *Psychology of Addictive Behaviors*, 26(3), 661–664. <https://doi.org/10.1037/a0027887>
- Khadjesari, Z., Vitoratou, S., Sevdalis, N., & Hull, L. (2017). Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: A systematic review protocol. *British Medical Journal Open*, 7(10), Article e017972. <https://doi.org/10.1136/bmjopen-2017-017972>
- Kirchner, J. E., Ritchie, M. J., Pitcock, J. A., Parker, L. E., Curran, G. M., & Fortney, J. C. (2014). Outcomes of a partnered facilitation strategy to implement primary care-mental health. *Journal of General Internal Medicine*, 29(Suppl. 4), 904–912. <https://doi.org/10.1007/s11606-014-3027-2>
- Kobak, K. A., Mundt, J. C., & Kennard, B. (2015). Integrating technology into cognitive behavior therapy for adolescent depression: A pilot study. *Annals of General Psychiatry*, 14, Article 37. <https://doi.org/10.1186/s12991-015-0077-8>
- Kobak, K. A., Stone, W. L., Ousley, O. Y., & Swanson, A. (2011). Web-based training in early autism screening: Results from a pilot study. *Telemedicine and E-Health*, 17(8), 640–644. <https://doi.org/10.1089/tmj.2011.0029>
- Kolko, D. J., Baumann, B. L., Herschell, A. D., Hart, J. A., Holden, E. A., & Wisniewski, S. R. (2012). Implementation of AF-CBT by community practitioners serving child welfare and mental health: A randomized trial. *Child Maltreatment*, 17(1), 32–46. <https://doi.org/10.1177/1077559511427346>
- Kolla, N. J., Links, P. S., McMain, S., Streiner, D. L., Cardish, R., & Cook, M. (2009). Demonstrating adherence to guidelines for the treatment of patients with borderline personality disorder. *The Canadian Journal of Psychiatry*, 54(3), 181–189. <https://doi.org/10.1177/070674370905400306>
- Kutsick, K. A., Witt, J. C., & Gutkin, T. B. (1991). The impact of treatment development process, intervention type, and problem severity on treatment acceptability as judged by classroom teachers. *Psychology in the Schools*, 28(4), 325–331. [https://doi.org/10.1002/1520-6807\(199110\)28:4<325::AID-PITS2310280407>3.0.CO;2-Y](https://doi.org/10.1002/1520-6807(199110)28:4<325::AID-PITS2310280407>3.0.CO;2-Y)
- Lehman, W. E. K., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, 22(4), 197–209. [https://doi.org/10.1016/S0740-5472\(02\)00233-7](https://doi.org/10.1016/S0740-5472(02)00233-7)
- Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*, 10, Article 155. <https://doi.org/10.1186/s13012-015-0342-x>
- Lewis, C. C., Mettert, K. D., Dorsey, C. N., Martinez, R. G., Weiner, B. J., Nolen, E., . . . Powell, B. J. (2018). An updated protocol for a systematic review of implementation-related measures. *Systematic Reviews*, 7(1), Article 66. <https://doi.org/10.1186/s13643-018-0728-3>
- Lloyd, D., Couineau, A. L., Hawkins, K., Kartal, D., Nixon, R. D., Perry, D., & Forbes, D. (2015). Preliminary outcomes of implementing cognitive processing therapy for posttraumatic stress disorder across a national veterans' treatment service. *The Journal of Clinical Psychiatry*, 76(11), e1405–e1409. <https://doi.org/10.4088/JCP.14m09139>
- Luke, D. A., Calhoun, A., Robichaux, C. B., Elliott, M. B., & Moreland-Russell, S. (2014). The Program Sustainability Assessment Tool: A new instrument for public health programs. *Preventing Chronic Disease*, 11, Article 130184. <https://doi.org/10.5888/pcd11.130184>
- Lyon, A. (2010). *Influences on Psychotherapy Training Participation Scale (IPTPS)*. [https://societyforimplementationresearchcollaboration.org/wp-content/uploads/2012/10/M\\_Lyon\\_IPTPS\\_2010.pdf](https://societyforimplementationresearchcollaboration.org/wp-content/uploads/2012/10/M_Lyon_IPTPS_2010.pdf)
- Lyon, A. (2011). *Training/Practice Acceptability/Feasibility/Appropriateness Scale* [Unpublished manuscript]. Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA.
- Maher, L., Gustafson, D., & Evans, A. (2007). *Sustainability and model guide*. NHS Institute for Innovation and Improvement.
- Malte, C. A., McFall, M., Chow, B., Beckham, J. C., Carmody, T. P., & Saxon, A. J. (2013). Survey of providers' attitudes toward integrating smoking cessation treatment into posttraumatic stress disorder care. *Psychology of Addictive Behaviors*, 27(1), 249–255. <https://doi.org/10.1037/a0028484>
- Mancini, J. A., & Marek, L. I. (2004). Sustaining community-based programs for families: Conceptualization and measurement. *Family Relations*, 53(4), 339–347. <https://doi.org/10.1111/j.0197-6664.2004.00040.x>
- Martinez, R. G., Lewis, C. C., & Weiner, B. J. (2014). Instrumentation issues in implementation science. *Implementation Science*, 9, Article 118. <https://doi.org/10.1186/s13012-014-0118-8>
- McIntosh, K. M. L. D., Hume, A. E., Doolittle, J., Vincent, C. G., Horner, R. H., & Ervin, R. A. (2011). Development and initial validation of a measure to assess factors related to sustainability of school-wide positive behavior support. *Journal of Positive Behavior Interventions*, 13(4), 208–218. <https://doi.org/10.1177/1098300710385348>
- McKee, S. A., Harris, G. T., & Cormier, C. A. (2013). Implementing residential integrated treatment for co-occurring disorders. *Journal of Dual Diagnosis*, 9(3), 249–259. <https://doi.org/10.1080/15504263.2013.807073>
- McLean, K. A. (2013). *Healthcare provider acceptability of a behavioral intervention to promote adherence* [Open Access database]. University of Miami. [https://scholarlyrepository.miami.edu/oa\\_theses/434](https://scholarlyrepository.miami.edu/oa_theses/434)
- 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. <https://doi.org/10.1037/pas0000037>
- Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. Guilford Press.
- Milne, D. (2010). Can we enhance the training of clinical supervisors? A national pilot study of an evidence-based approach. *Clinical Psychology & Psychotherapy*, 17(4), 321–328. <https://doi.org/10.1002/cpp.657>
- Milne, D., & Noone, S. (1996). *Teaching and training for non-teachers*. The British Psychological Society.

- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192–222. <https://doi.org/10.1287/isre.2.3.192>
- Myers, S. J. (2008). *Relationship between the consultant-parent working alliance and ratings of the consultation process with parents of children having autism spectrum disorder* [Doctoral dissertation]. The University of Arizona.
- Neumiller, S., Bennett-Clark, F., Young, M. S., Dates, B., Broner, N., Leddy, J., & De Jong, F. (2009). Implementing assertive community treatment in diverse settings for people who are homeless with co-occurring mental and addictive disorders: A series of case studies. *Journal of Dual Diagnosis*, 5(3–4), 239–263. <https://doi.org/10.1080/15504260903175973>
- Newton, J. T., & Sturme, P. (2004). Development of a short form of the Treatment Evaluation Inventory for acceptability of psychological interventions. *Psychological Reports*, 94(2), 475–481. <https://doi.org/10.2466/pr0.94.2.475-481>
- Oxman, T. E., Schulberg, H. C., Greenberg, R. L., Dietrich, A. J., Williams, J. W., Jr., Nutting, P. A., & Bruce, M. L. (2006). A fidelity measure for integrated management of depression in primary care. *Medical Care*, 44(11), 1030–1037. <https://doi.org/10.1097/01.mlr.0000233683.82254.63>
- Pace, C. A., Buczek, J., Cruz, A., Ellenberg, L., Jonathan, F., & Ginman, E. W. (2014). Integrating behavioral health services and primary care at an urban, safety-net teaching hospital: A pilot program. *Journal of General Internal Medicine*, 23, S487–S488.
- Pankratz, M., Hallfors, D., & Cho, H. (2002). Measuring perceptions of innovation adoption: The diffusion of a federal drug prevention policy. *Health Education Research*, 17(3), 315–326. <https://doi.org/10.1093/her/17.3.315>
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829–841. <https://doi.org/10.1037/0022-006X.75.6.829>
- Power, T. J., Hess, L. E., & Bennett, D. S. (1995). The acceptability of interventions for attention-deficit hyperactivity disorder among elementary and middle school teachers. *Journal of Developmental and Behavioral Pediatrics*, 16(4), 238–243. <https://doi.org/10.1097/00004703-199508000-00005>
- Proctor, E., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research*, 36(1), 24–34. <https://doi.org/10.1007/s10488-008-0197-4>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . 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. <https://doi.org/10.1007/s10488-010-0319-7>
- Reimers, T. M., & Wacker, D. P. (1988). Parents' ratings of the acceptability of behavioral treatment recommendations made in an outpatient clinic: A preliminary analysis of the influence of treatment effectiveness. *Behavioral Disorders*, 14(1), 7–15. <https://doi.org/10.1177/019874298801400104>
- Roundfield, K. D., & Lang, J. M. (2017). Costs to community mental health agencies to sustain an evidence-based practice. *Psychiatric Services*, 68(9), 876–882. <https://doi.org/10.1176/appi.ps.201600193>
- Ruzek, J. I., Eftekhari, A., Rosen, C. S., Crowley, J. J., Kuhn, E., Foa, E. B., . . . Karlin, B. E. (2016). Effects of a comprehensive training program on clinician beliefs about and intention to use prolonged exposure therapy for PTSD. *Psychological Trauma: Theory, Research, Practice, and Policy*, 8(3), 348–355. <https://doi.org/10.1037/tra0000004>
- Simpson, D. D. (2002). A conceptual framework for transferring research to practice. *Journal of Substance Abuse Treatment*, 22(4), 171–182. [https://doi.org/10.1016/s0740-5472\(02\)00231-3](https://doi.org/10.1016/s0740-5472(02)00231-3)
- Skinner, K. (2007). Developing a tool to measure knowledge exchange outcomes. *The Canadian Journal of Program Evaluation*, 22(1), 49–73.
- Slade, M., Cahill, S., Kelsey, W., Powell, R., Strathdee, G., & Valiakalayil, A. (2001). Threshold 3: The feasibility of the Threshold Assessment Grid (TAG) for routine assessment of the severity of mental health problems. *Social Psychiatry and Psychiatric Epidemiology*, 36(10), 516–521. <https://doi.org/10.1007/s001270170017>
- Solberg, L. I., Asche, S. E., Margolis, K. L., & Whitebird, R. R. (2008). Measuring an organization's ability to manage change: The Change Process Capability Questionnaire and its use for improving depression care. *American Journal of Medical Quality*, 23(3), 193–200. <https://doi.org/10.1177/1062860608314942>
- Stamatakis, K. A., McQueen, A., Filler, C., Boland, E., Dreisinger, M., Brownson, R. C., & Luke, D. A. (2012). Measurement properties of a novel survey to assess stages of organizational readiness for evidence-based interventions in community chronic disease prevention settings. *Implementation Science*, 7, Article 65. <https://doi.org/10.1186/1748-5908-7-65>
- Stanick, C. F., Halko, H. M., Nolen, E. A., Powell, B. J., Dorsey, C. N., Mettert, K. D., . . . Lewis, C. C. (2019). Pragmatic measures for implementation research: Development of the Psychometric and Pragmatic Evidence Rating Scale (PAPERS). *Translational Behavioral Medicine*. <https://doi.org/10.1093/tbm/ibz164>
- Swain, K., Whitley, R., McHugo, G. J., & Drake, R. E. (2010). The sustainability of evidence-based practices in routine mental health agencies. *Community Mental Health Journal*, 46(2), 119–129. <https://doi.org/10.1007/s10597-009-9202-y>
- Tanol, G. (2010). *Treatment fidelity: Relation to treatment acceptability and change over time*. [https://conservancy.umn.edu/bitstream/handle/11299/101133/Tanol\\_umn\\_0130E\\_11548.pdf?sequence=1&isAllowed=y](https://conservancy.umn.edu/bitstream/handle/11299/101133/Tanol_umn_0130E_11548.pdf?sequence=1&isAllowed=y)
- Tanol, G., Johnson, L., McComas, J., & Cote, E. (2010). Responding to rule violations or rule following: A comparison of two versions of the Good Behavior Game with kindergarten students. *Journal of School Psychology*, 48(5), 337–355. <https://doi.org/10.1016/j.jsp.2010.06.001>
- Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry*, 23(2), 101–106. [https://doi.org/10.1016/0005-7916\(92\)90007-6](https://doi.org/10.1016/0005-7916(92)90007-6)
- Tobin, M. J., Clarke, A. R., Buss, R., Einfeld, S. L., Beard, J., Dudley, M., . . . Dietrich, U. (2001). From efficacy to

effectiveness: Managing organisational change to improve health services for young people with deliberate self harm behaviour. *Australian Health Review*, 24(2), 143–151. <https://doi.org/10.1071/ah010143>

Trent, L. R. (2010). *Development of a measure of disseminability (MOD)*. University of Mississippi.

Walitzer, K. S., Dermen, K. H., Barrick, C., & Shyhalla, K. (2015). Modeling the innovation-decision process: Dissemination and adoption of a motivational interviewing preparatory procedure in addiction outpatient clinics. *Journal of Substance Abuse Treatment*, 57, 18–29. <https://doi.org/10.1016/j.jsat.2015.04.003>

Weeks, G., Slade, M., & Hayward, M. (2011). A UK validation of the stages of recovery instrument. *International Journal of Social Psychiatry*, 57(5), 446–454. <https://doi.org/10.1177/0020764010365414>

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child & Adolescent Psychology*, 31(2), 168–180. [https://doi.org/10.1207/S15374424JCCP3102\\_03](https://doi.org/10.1207/S15374424JCCP3102_03)

Whittingham, K., Sofronoff, K., & Sheffield, J. K. (2006). Stepping Stones Triple P: A pilot study to evaluate acceptability of the program by parents of a child diagnosed with an autism spectrum disorder. *Research in Developmental Disabilities*, 27(4), 364–380. <https://doi.org/10.1016/j.ridd.2005.05.003>

Williams, N. J., Glisson, C., Hemmelgarn, A., & Green, P. (2017). Mechanisms of change in the ARC organizational strategy: Increasing mental health clinicians’ EBP adoption through improved organizational culture and capacity. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(2), 269–283. <https://doi.org/10.1007/s10488-016-0742-5>

Woltmann, E. M., Whitley, R., McHugo, G. J., Brunette, M., Torrey, W. C., Coats, L., . . . Drake, R. E. (2008). The role of staff turnover in the implementation of evidence-based practices in mental health care. *Psychiatric Services*, 59(7), 732–737. <https://doi.org/10.1176/appi.ps.59.7.732>

Yetter, G. (2010). Assessing the acceptability of problem-solving procedures by school teams: Preliminary development of the pre-referral intervention team inventory. *Journal of Educational and Psychological Consultation*, 20(2), 139–168. <https://doi.org/10.1080/10474411003785370>

Zwerver, F., Schellart, A. J., Anema, J. R., & van der Beek, A. J. (2013). Changes in insurance physicians’ attitudes, self-efficacy, intention, and knowledge and skills regarding the guidelines for depression, following an implementation strategy. *Journal of Occupational Rehabilitation*, 23(1), 148–156. <https://doi.org/10.1007/s10926-012-9378-9>

## Appendix I

### PRISMA flowcharts for included and excluded studies

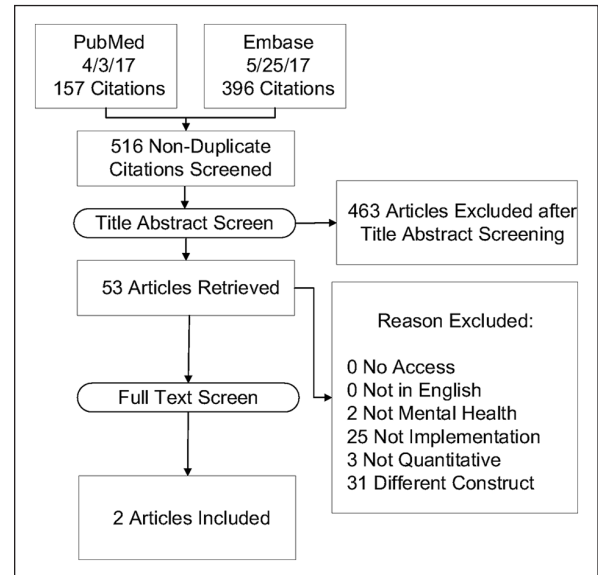


Figure A1. Acceptability PRISMA flowchart.

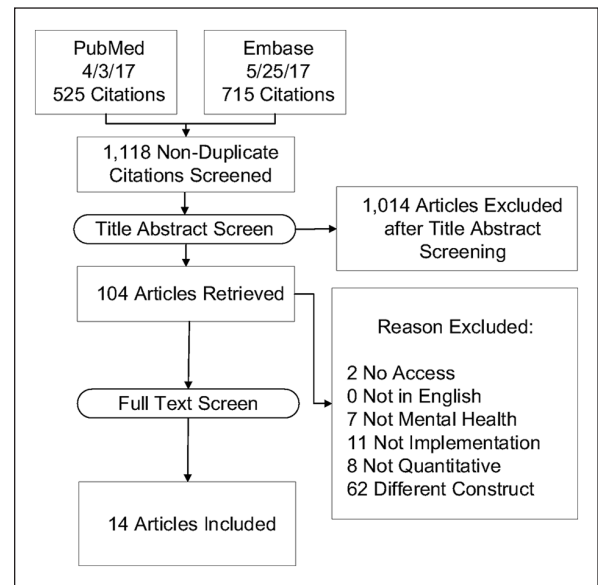


Figure A2. Adoption PRISMA flowchart.

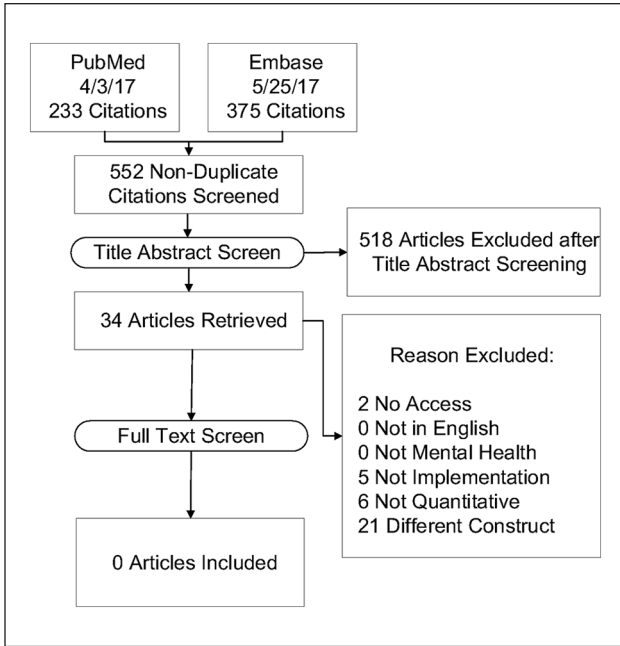


Figure A3. Appropriateness PRISMA flowchart.

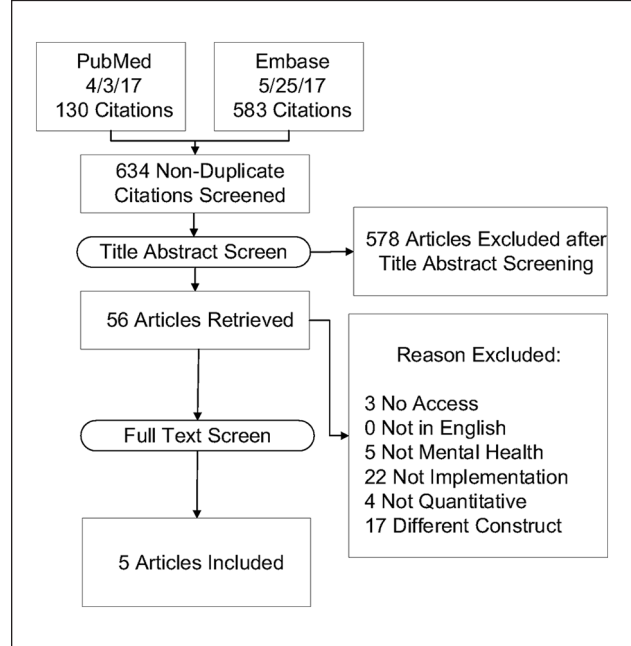


Figure A5. Feasibility PRISMA flowchart.

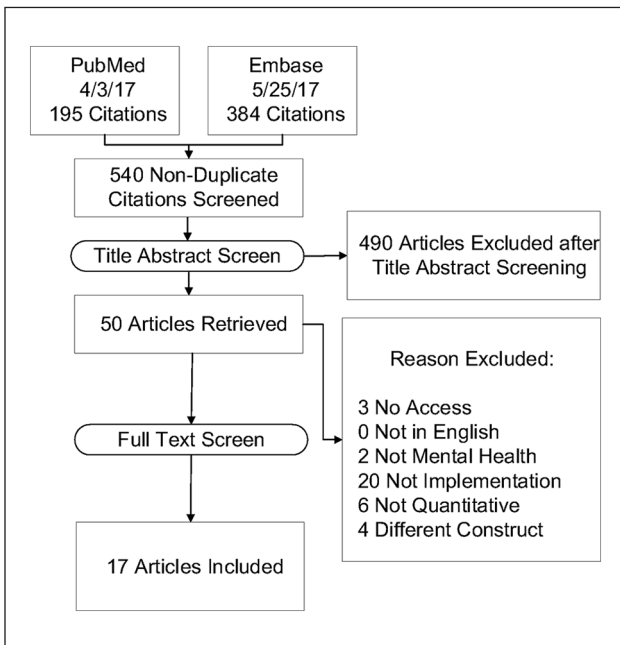


Figure A4. Cost PRISMA flowchart.

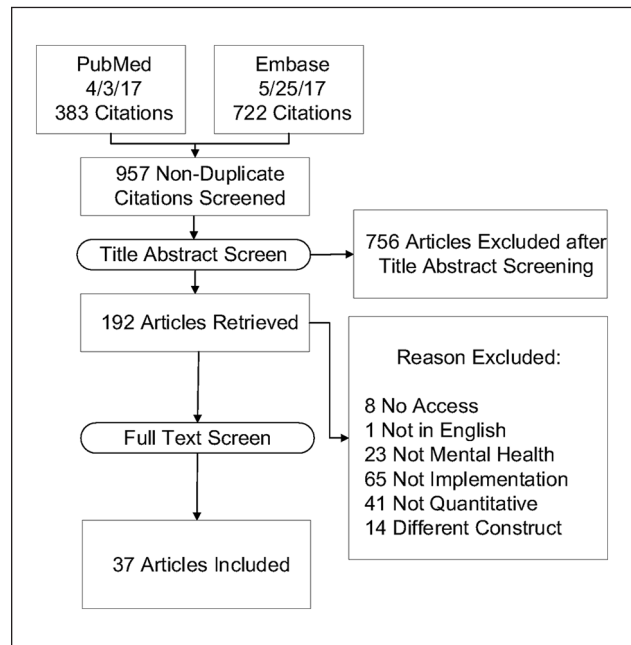
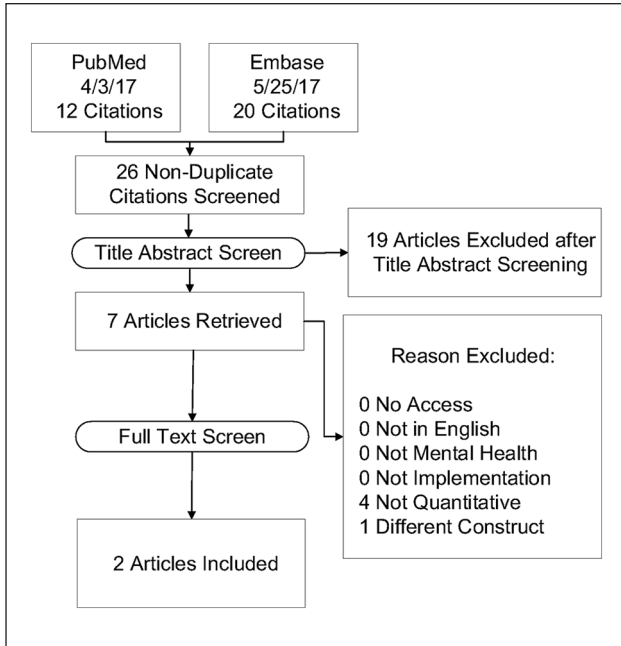
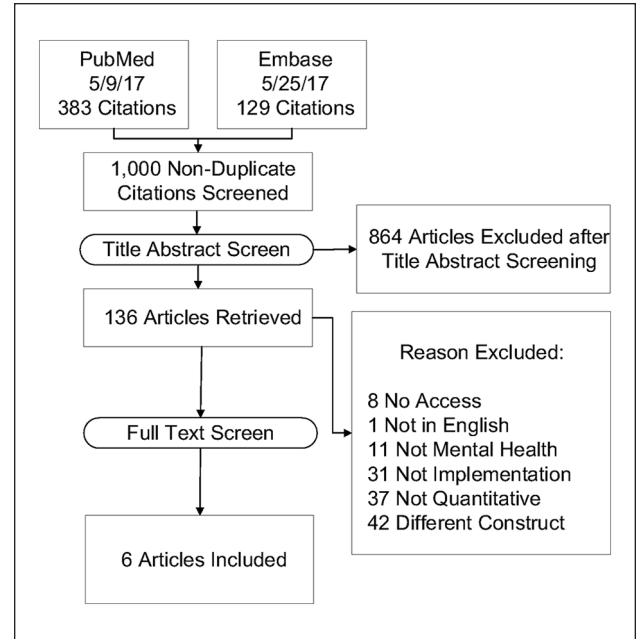


Figure A6. Fidelity PRISMA flowchart.



**Figure A7.** Penetration PRISMA flowchart.



**Figure A8.** Sustainability PRISMA flowchart.