

Title	Measures of outer setting constructs for implementation research: a systematic review and analysis of psychometric quality
Authors	McHugh, Sheena M.;Dorsey, Caitlin N.;Mettert, Kayne;Purtle, Jonathan;Bruns, Eric;Lewis, Cara C.
Publication date	2020-01
Original Citation	McHugh, S. M., Dorsey, C. N., Mettert, K., Purtle, J., Bruns, E. and Lewis, C. C. (2020) 'Measures of outer setting constructs for implementation research: a systematic review and analysis of psychometric quality'. Implementation Research and Practice, 1 , [20pp]. doi: 10.1177/2633489520940022
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1177/2633489520940022
Rights	© The Author(s) 2020. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://uk.sagepub.com/aboutus/openaccess.htm). - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2025-03-18 00:59:06
Item downloaded from	https://hdl.handle.net/10468/13158



UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



Measures of outer setting constructs for implementation research: A systematic review and analysis of psychometric quality

Implementation Research and Practice
Volume 1: Jan-Dec 2020 1–20
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2633489520940022
journals.sagepub.com/home/irp

Sheena McHugh¹ , Caitlin N Dorsey², Kayne Mettert² ,
Jonathan Purtle³, Eric Bruns⁴ and Cara C Lewis²

Abstract

Background: Despite their influence, outer setting barriers (e.g., policies, financing) are an infrequent focus of implementation research. The objective of this systematic review was to identify and assess the psychometric properties of measures of outer setting used in behavioral and mental health research.

Methods: Data collection involved (a) search string generation, (b) title and abstract screening, (c) full-text review, (d) construct mapping, and (e) measure forward searches. Outer setting constructs were defined using the Consolidated Framework for Implementation Research (CFIR). The search strategy included four relevant constructs separately: (a) cosmopolitanism, (b) external policy and incentives, (c) patient needs and resources, and (d) peer pressure. Information was coded using nine psychometric criteria: (a) internal consistency, (b) convergent validity, (c) discriminant validity, (d) known-groups validity, (e) predictive validity, (f) concurrent validity, (g) structural validity, (h) responsiveness, and (i) norms. Frequencies were calculated to summarize the availability of psychometric information. Information quality was rated using a 5-point scale and a final median score was calculated for each measure.

Results: Systematic searches yielded 20 measures: four measures of the general outer setting domain, seven of cosmopolitanism, four of external policy and incentives, four of patient needs and resources, and one measure of peer pressure. Most were subscales within full scales assessing implementation context. Typically, scales or subscales did not have any psychometric information available. Where information was available, the quality was most often rated as “1-minimal” or “2-adequate.”

Conclusion: To our knowledge, this is the first systematic review to focus exclusively on measures of outer setting factors used in behavioral and mental health research and comprehensively assess a range of psychometric criteria. The results highlight the limited quantity and quality of measures at this level. Researchers should not assume “one size fits all” when measuring outer setting constructs. Some outer setting constructs may be more appropriately and efficiently assessed using objective indices or administrative data reflective of the system rather than the individual.

Keywords

Implementation science, systematic review, measures, scale, psychometric, incentive, policy

¹School of Public Health, University College Cork, Cork, Ireland

²Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

³Department of Health Management & Policy, Dornsife School of Public Health, Drexel University, Philadelphia, PA, USA

⁴Department of Psychiatry and Behavioural Sciences, University of Washington School of Medicine, Seattle, WA, USA

Corresponding author:

Sheena McHugh, School of Public Health, University College Cork, Western Gateway Building, Western Rd, Cork, T12 XF62 Ireland.
Email: s.mchugh@ucc.ie



Plain language abstract

Implementation of evidence-based practices (EBPs) is influenced by the wider environment or *outer setting* in which it takes place. Despite their influence, barriers in the outer setting, such as policies, financing, and stakeholder relationships, are often not the focus of implementation research. The aim of this systematic review was to identify and evaluate measures used in behavioral and mental health research to assess barriers in the outer setting. We collected and analyzed all relevant implementation studies in settings such as mental health services, psychiatry, and substance use. We looked for self-reported measures of cosmopolitanism (connections between organizations), policies and incentives, patient needs and resources, and peer pressure. We identified 20 measures of outer setting factors. Most measures were subsections of longer tools assessing a range of implementation barriers. Most measures did not have any information available on their validity and reliability. Where this information was available, the quality of the information was rated as minimal or adequate. While other reviews have examined barriers to implementation at different levels in different settings, this is the first systematic review to focus exclusively on measures of outer setting factors used in behavioral health research. The results highlight the limited quantity and quality of measurement tools. It may be faster and more appropriate to assess some outer setting factors using data already collected on administrative systems. The results of this systematic review will guide researchers and practitioners on the use and testing of measures in future evaluations.

Implementation is a social process that is inherently dependent on the context in which it takes place (Davidoff et al., 2008). In implementation research, context refers to the unique set of circumstances that surround a particular implementation effort (Damschroder et al., 2009). There are a plethora of implementation models and frameworks that categorize context into groups of factors that are hypothesized to affect implementation outcomes (Tabak et al., 2012). Most frameworks distinguish between contextual factors related to the organizational setting in which implementation is taking place and factors related to the wider environment in which the organization operates (Aarons et al., 2011; Damschroder et al., 2009; Feldstein & Glasgow, 2008). The Consolidated Framework for Implementation Research (CFIR) refers to these groups of contextual factors as the inner and outer setting, respectively (Damschroder et al., 2009). Within the outer setting, an array of social, fiscal, and policy factors are considered important for implementation success. However, most frameworks have not been tested empirically to assess their utility to explain the influence of contextual factors, including outer setting factors, on implementation outcomes. Robust quantitative measures are needed to identify causal predictors of implementation outcomes.

Outer setting factors are notoriously difficult to evaluate and influence, and efforts to operationalize constructs within this domain are scant. There are several reasons for this paucity of research on the outer setting generally and the development of measures specifically, most of which are borne out of methodological challenges that face potential investigators. First, outer setting factors such as policy and fiscal issues are hard to experimentally manipulate in research studies which may result in fewer studies attempting to measure their independent influence on implementation outcomes. Second, unlike individuals or organizations (two other primary levels of implementation), the policy and funding context is less easily operationalized and isolated as a unit of measurement. Third, similar to other potentially influential determinants of implementation, there are no explicit recommendations for

operational definitions or items to measure outer setting constructs (Cook et al., 2012). These and other methodological challenges have led researchers to afford relatively less attention to outer setting factors, even while citing the criticality of these factors to successful implementation.

To date, several systematic reviews of implementation measures have been conducted, many of which have concluded that there is a lack of reliable and valid measures available (Chaudoir et al., 2013; Chor et al., 2015; Emmons et al., 2012; Weiner et al., 2008). Results from these studies also highlight the gap in the range of contextual factors for which measures are currently available. A number of reviews have concentrated solely on the measurement of inner setting constructs (Emmons et al., 2012; French et al., 2009; Weiner et al., 2008). In systematic reviews that have examined the full range of contextual factors (Chaudoir et al., 2013; Chor et al., 2015; Clinton-McHarg et al., 2016; Kaplan et al., 2010), few measures of outer setting constructs have been identified. A systematic review of business and health care literature examining the influence of context on the success of quality improvement (QI) efforts in health care settings (units within hospitals, hospitals, and integrated delivery systems) identified 15 (out of 47) studies examining the influence of “environment.” There was inconsistent evidence of an association between these factors and QI success (Kaplan et al., 2010). The review did not examine the psychometric properties of measures used to assess these environmental factors, but if the quality of these measures is similar to those assessed in related reviews (i.e., poor quality), then it may be that measurement issues undermined any ability to detect a relation.

A systematic review of factors affecting implementation of health-related innovations identified only five measures of “structural-level” constructs (8% of all measures; 62 measures in total). While the review included studies in a range of settings including mental health settings (as well as educational, work place, and other settings), all five measures of “structural-level” constructs were applied in health care settings (Chaudoir et al., 2013). Of the five measures, four were assessed for criterion (predictive) validity. Of those measures,

three were statistically significant predictors of adoption. A recent review of measures used in public health and community settings, which used CFIR as its organizing framework, identified 18 measures of outer context out of a total of 51 implementation measures (35%) (Clinton-McHarg et al., 2016). This review focused on non-clinical settings where the delivery of health or mental health was not the primary focus. Overall, psychometric information was poorly reported and where available, measures demonstrated limited reliability and validity. Finally, in a review of over 100 measures of predictors of the adoption of EBPs in mental health specifically, only 18 measures assessed the “external system,” six of which had any evidence of reliability or validity (Chor et al., 2015). Consistent, accurate, and valid measures of all levels of context are crucial to enable a comprehensive understanding of factors influencing implementation outcomes.

While some reviews have included similar implementation contexts such as correctional facilities (Clinton-McHarg et al., 2016) and mental health/substance abuse settings (Chaudoir et al., 2013), to our knowledge, no previous systematic reviews of measures of implementation constructs have focused exclusively on *outer setting constructs* for use in behavioral and mental health research and comprehensively assessed a range of psychometric criteria. Behavioral and mental health research was defined as that which concerned mental health, substance use, or other addictive behaviors. The study of outer setting determinants is particularly important in behavioral and mental health. Responsibility for behavioral and mental health care is diffuse (Purtle et al., 2017). It relies on complex interactions between multiple agencies, leaders, funders, and policy makers, across national, state, and local levels, all of which vary greatly across government entities. This complexity is coupled with public and policymaker stigma toward people with mental illness (Purtle et al., 2017; Smith et al., 2012). The focus of this review on behavioral and mental health was also consistent with the priorities of the Society for Implementation Research Collaboration (SIRC) and our funding agency, the National Institute of Mental Health (Lewis, Mettert, et al 2018). In this systematic review, we focused on implementation constructs from the outer setting domain as defined by the CFIR (Damschroder et al., 2009): (a) cosmopolitanism, (b) external policies and incentives, (c) patient needs and resources, and (d) peer pressure. The aim of this systematic review was to identify measures of outer setting and its CFIR-delineated constructs used in behavioral and mental health research, and assess the psychometric properties of those measures.

Method

Design overview

Data for this study come from a larger project funded by the National Institute of Mental Health—the SIRC Instrument

Review Project—which aims to advance implementation science by identifying quantitative measures used in behavioral and mental health that demonstrate both psychometric and pragmatic strength. Full details of the protocol for the entire set of systematic reviews have been published elsewhere (Lewis, Mettert, et al 2018). Systematic review methodology was chosen as we wanted to go beyond identifying available measures, to establish and rate the quality of psychometric evidence using explicit systematic methods. Consistent with the larger project, measure identification and evaluation consisted of three phases. Phase I, measure identification, included the following five steps: (a) search string generation, (b) title and abstract screening, (c) full-text review, (d) measure mapping to CFIR construct, and (e) measure forward (cited-by) searches. Phase II, data extraction, consisted of coding relevant psychometric information. In Phase III, data analysis was completed.

Phase I: data collection

First, systematic searches were conducted in PubMed and Embase bibliographic databases using search strings curated in consultation with PubMed support specialists and a library scientist. Consistent with our aim to identify and assess implementation-related measures in the behavioral and mental health space, our search was built on four core levels: (a) terms for implementation, (b) terms for EBP, (c) terms for measurement, and (d) terms for behavioral and mental health. Table 1 presents the search terms used. We included a fifth level of terms for each of the following outer setting constructs from CFIR: (a) external policy and incentives, (b) cosmopolitanism, (c) patient needs and resources, and (d) peer pressure. In CFIR, cosmopolitanism is defined as “the degree to which an organization is networked with other external organizations.” External policies and incentives refer to “external strategies to spread interventions, including policy and regulations (governmental or other central entity), external mandates, recommendations and guidelines, pay-for-performance, collaboratives, and public or benchmark reporting.” Patient needs and resources is defined as

the extent to which patient needs, as well as barriers and facilitators to meet those needs, are accurately known and prioritized by the organization while peer pressure refers to mimetic or competitive pressure to implement an intervention; typically because most or other key peer or competing organizations have already implemented or are in a bid for a competitive edge. (Damschroder et al., 2009)

Systematic searches were conducted independently for each construct; four different search strings were used. Articles published from 1985 onward were included in the search. Electronic database searches were completed up to May 2017.

Identified articles were subject to title and abstract screening followed by full-text review, to confirm relevance to the study parameters. In brief, we included only empirical studies that contained one or more quantitative measures of the outer setting domain or its four constructs if they were used in an evaluation of an implementation effort in a behavioral health context; see Table 2 for a breakdown of the inclusion/exclusion criteria. We also included additional measures of outer setting constructs that were identified during screening of articles relating to the other CFIR constructs evaluated in the larger project (e.g., screening articles on inner setting measures produced measures relevant to outer setting).

Included articles then progressed to the fourth step, construct mapping. Trained research specialists (C.N.D., K.M.) mapped measures and/or their scales to the outer setting domain and/or one or more of the four aforementioned CFIR constructs using a two-pronged approach. First, research specialists mapped measures to CFIR constructs based on the study author's definition or description of which construct was purportedly being measured. To account for additional content areas not captured by the description or in the absence of a clear definition, the research specialists reviewed each measure's item pool and mapped the measure to an outer setting construct if two or more items were identified as assessing that construct (Lewis, Mettert, et al., 2018). If items were not provided for coding, research specialists mapped the measure more broadly to the outer setting domain. Construct assignment was checked and confirmed by content experts (SMcH, J.P., E.B.) having reviewed items within each measure and/or scale.

In the final step, each included measure was subjected 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," which included the measure itself (as available), the measurement development article (or article with the first empirical use in a behavioral health context), and all additional empirical uses of the measure in behavioral health. To identify all relevant reports of psychometric information, the team of trained research specialists (C.N.D., K.M.) reviewed each article and electronically extracted information to assess the psychometric rating criteria, using a rating system with relevant criteria referred to hereafter as PAPERS (Psychometric And Pragmatic Evidence Rating Scale). The full rating system and criteria for the PAPERS is published elsewhere (Lewis, Mettert, et al., 2018). This study, which focuses on psychometric properties only, used only nine of the 14 PAPERS criteria: (a) internal consistency, (b) convergent validity, (c) discriminant

validity, (d) known-groups validity, (e) predictive validity, (f) concurrent validity, (g) structural validity, (h) responsiveness, and (i) norms. Data on each psychometric criterion were extracted for both full measure and individual scale levels as appropriate. Measures were considered "unsuitable for rating" if the format of construct assessment (a) would not allow for psychometric information to be produced (e.g., qualitative nomination form) or (b) 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: "−1-poor," "0-none," "1-minimal/emerging," "2-adequate," "3-good," or "4-excellent." 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 psychometric 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 peer pressure 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 extracted on each measure. The characteristics described the use of a measure in the behavioral health literature overall and was not specific to its use as a measure of outer setting constructs. Characteristics included (a) country of origin, (b) concept defined by authors, (c) number of articles contained in each measure packet, (d) number of scales, (e) number of items, (f) setting in which measure had been used, (g) level of analysis, (h) target problem, and (i) stage of implementation as defined by the Exploration, Adoption/Preparation,

Implementation, Sustainment (EPIS) model (Aarons et al., 2011). Where a measure did not have a formal name, the first author surname and general description was used to name measures.

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 allow for visual comparisons across all measures within a given construct.

Results

Overview of measures

Similar to other systematic reviews of implementation measures (Lewis et al., 2015), traditional systematic review methods were not useful for identifying articles with measures of outer setting constructs. Searches of two electronic bibliographic databases, PubMed and Embase, yielded 2,347 non-duplicate articles. Only five articles were retained following the review process (see Figures A1 to A5 in Appendix 1 for Construct PRISMA Flowcharts). Additional articles were identified during screening for other CFIR constructs and “cited by” searches, resulting in 57 articles being included in measurement packets for analysis. The number of articles identified did not equal the number of measures included in the analysis as multiple articles used the same measure and some articles used more than one measure.

Overall, systematic searches yielded 20 measures related in full or in part to the outer setting domain and/or its constructs, which had been used in behavioral health. As Table 3 illustrates, three measures mapped to more than one construct in the outer setting domain. In such cases, the measure was counted separately for each construct. One measure, a measure of “environmental dynamism” (Nieboer & Strating, 2012), which was part of a larger survey assessing the influence of organizational characteristics on innovation culture, was counted as an individual scale rather than a subscale. The larger survey contained several distinct validated scales which were relevant to different individual constructs: outer setting, cosmopolitanism, and peer pressure. As it was not clear whether these scales were part of one single validated measure, they were treated as separate scales in this analysis.

Four measures were identified for the general outer setting domain. There were three full scales related to outer setting and one subscale. In terms of individual constructs, seven measures of cosmopolitanism were identified. In six

cases, these were subscales within broader measures. In total, four measures of external policy and incentives were identified: two were full scales and two were subscales. Four measures of patient needs and resources were identified, all of which were subscale(s). Finally, one measure assessing peer pressure was identified.

Characteristics of measures

Table 4 presents the descriptive characteristics of all 20 measures which were used, in full or in part, to assess the outer setting domain and its constructs. In the behavioral health research literature, most measures were used only in a single evaluation or research study; in the United States, at the provider, supervisor, or director level; and in an outpatient community setting and a variety of “other” settings (e.g., prison, church).

Availability and rating of psychometric evidence

Of the 20 measures of outer setting and/or one of its constructs, four were categorized as unsuitable for rating. Table 5 presents a summary of the number of measures for which psychometric information was available. Table 6 describes the median ratings and range of ratings of psychometric properties for those measures deemed suitable for rating ($n=16$) and those for which information was available (i.e., those with non-zero ratings on psychometric criteria). Following the rolled-up approach applied in this study, results are presented at the level of full scale. Where appropriate, we highlight the number of subscales relevant to an outer setting construct within that scale.

Outer setting. Four measures of the general outer setting domain were identified in behavioral health research, three of which were suitable for rating. Three of these measures had psychometric information available. There was information on internal consistency, predictive validity, and norms for the two measures, and information on concurrent validity for one measure. There was no information available for convergent validity, discriminant validity, known-groups validity, structural validity, and responsiveness.

For measures of outer setting, the median rating was “2-adequate” for internal consistency, predictive validity, and norms. The median rating was “1-minimal” for concurrent validity. However, this was based on just one measure: Nieboer and Stratling’s environmental dynamism scale (Nieboer & Strating, 2012). This measure had the highest psychometric rating score of the three outer setting measures with information available (psychometric total score = 8 out of a maximum possible score of 36) with ratings of “3-good” for internal consistency and “2-adequate” for predictive validity and norms. However, its concurrent validity was rated “1-minimal,” and it had not been assessed for convergent validity, discriminant validity,

Table 1. Electronic bibliographic database search terms.

Level	PubMed search terms	Embase search terms
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 diffusion[tiab] OR "diffusion of innovation"[MeSH Terms] OR "health information exchange"[MeSH Terms] OR "knowledge translation"[tw] OR "knowledge exchange"[tw])	adopt:ab,ti OR adopts:ab,ti OR adopted:ab,ti OR adopting:ab,ti OR adoption:ab,ti NOT "adoption"/exp OR implement:ab,ti OR implements:ab,ti OR implementation:ab,ti OR "health care planning"/exp OR "quality improvement":ab,ti OR "total quality management"/exp OR diffused:ab,ti OR diffusion:ab,ti OR "mass communication"/exp OR "medical information system"/exp OR "knowledge translation":ab,ti OR "knowledge exchange":ab,ti
Evidence-based practice	AND ("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 guideline[tiab] OR guideline development[tiab] OR guideline[tiab] OR guidelineitem[tiab] OR guidelineon[tiab] OR guideline[tiab] OR guideline[tiab] OR guideline recommended[tiab] OR guideline related[tiab] OR guideline trade[tiab] OR guideline s[tiab] OR guideline s[tiab] OR guideline s quality[tiab] OR guideline s[tiab] OR guideline s[tiab] OR guideline s 9[tiab] OR guideline s 20[tiab] OR guideline s 20 female[tiab] OR guideline s for[tiab] OR guideline s in[tiab] OR guideline s may[tiab] OR guideline s on[tiab] OR guideline s[tiab] OR guideline s that[tiab] OR guideline s trade[tiab] OR guideline s wiki[tiab] OR "guidelines as topic"[MeSH Terms] OR "best practice"[tw])	AND "empirically supported treatment" OR "evidence based practice" OR "evidence based treatment" OR "evidence based practice" OR "evidence based treatment" OR innovation OR "practice guideline"/exp OR guideline:ab,ti [NOTE: EBP contains EBM]

(Continued)

Table 1. (Continued)

Level	PubMed search terms	Embase search terms
Measurement	<p>AND (instrument[tw] OR survey[tw] OR survey's[tw] OR survey100[tw] OR survey12[tw] OR survey1988[tw] OR survey226[tw] OR survey36[tw] OR surveyability[tw] OR surveyable[tw] OR surveyance[tw] OR surveysans[tw] OR surveyansin[tw] OR surveybetween[tw] OR surveyd[tw] OR surveydagger[tw] OR surveydata[tw] OR surveydelin[tw] OR surveyed[tw] OR surveydandtestedthe[tw] OR surveyedpopulation[tw] OR surveyees[tw] OR surveyelicited[tw] OR surveyer[tw] OR surveys[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'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 surveytas[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 \$[tw] OR questionnaire [tw] OR questionnaire [tw] OR questionnaire 12[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 questionnaire[tw] OR questionnairedeveloped[tw] OR questionnaireepq[tw] OR questionnaireforepediatric[tw] OR questionnairegr[tw] OR questionnairehas[tw] OR questionnaireitq[tw] OR questionnaire02[tw] OR questionnaireremcesqscale[tw] OR questionnaireurse[tw] OR questionnaireon[tw] OR questionnaireonline[tw] OR questionnairerep[tw] OR questionnairerephq[tw] OR questionnairers[tw] OR questionnaires[tw] OR questionnaires'[tw] OR questionnaires" [tw] OR questionnaire\$can[tw] OR questionnaire\$dq adolescent [tw] OR questionnaire\$ [tw] OR questionnairetrade[tw] OR questionnaireure[tw] OR questionnairev [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 \$[tiab] OR measurement [tiab] OR measuremental[tiab] OR measurement[tiab] OR measurement[tiab] OR measurementexhaled[tiab] OR measurement[tiab] OR measurementin[tiab] OR measurementin[tiab] OR measurementis[tiab] OR measurementcomputation[tiab] OR measurement[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 \$[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 measurements[tiab] OR measures[tiab] OR inventory[tiab])</p>	<p>AND instrument*:ab,ti OR survey*:ab,ti OR questionnaire*:ab,ti OR "questionnaire"/exp OR measurement*:ab,ti OR measure*:ab,ti OR inventory:ab,ti</p>

(Continued)

Table 1. (Continued)

Level	PubMed search terms	Embase search terms
Behavioral health	AND ("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 (psychiatrists[tw] OR psychiatrist[s] OR psychiatrists[tw] OR psychiatrists[tw] OR psychiatrists[quality] 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 "mental health":ab,ti OR "behavioral health":ab,ti OR psychiatr:ab,ti OR "mental disease"/exp OR "psychiatry"/exp OR "behavioral medicine"/exp OR "mental health service"/exp OR "mental hospital"/exp OR "psychiatric nursing"/exp
Patient needs and resources	"patient needs" OR "patient resources" OR "patient barriers" OR "patient satisfaction" OR "patient facilitators"	"patient needs" OR "patient resources" OR "patient choices" OR "patient barriers" OR "patient facilitators"
Cosmopolitanism	cosmopolitanism [tw] OR network [tw] OR "social capital" OR "external bridging" OR cosmopolitanism	cosmopolitanism OR "social capital" OR "external bridging"
Peer pressure	"mimetic isomorphism" OR "peer pressure" OR "competitive pressure" OR "pressure to implement"	"mimetic isomorphism" OR "peer pressure" OR "competitive pressure" OR "pressure to implement"
External policy and incentives	policy [tw] OR policies[tw] OR health policy[mh] OR regulations [tw] OR "external mandat*" OR guideline* OR recommendation OR "pay-for-performance" OR collaborative [tw] OR "public reporting" [tw] OR "benchmark reporting" OR guideline* [tw] OR guidelines as topic[mh] OR guideline[pt] OR recommendation*[tw]	policy OR policies OR "health care policy"/exp OR regulations OR "external mandat*" OR guideline* OR recommendation* OR "pay-for-performance" OR collaborative OR "public reporting" OR "benchmark reporting" OR "practice guideline"/exp

Table 2. Inclusion and exclusion criteria.

Level	Inclusion criteria	Exclusion criteria
Intervention	<ul style="list-style-type: none"> Behavioral health interventions broadly construed, typically these are psychosocial interventions (e.g., cognitive behavioral therapy, motivational interviewing, multisystemic therapy) Behavioral health interventions could also include care coordination, case management, and screening 	<ul style="list-style-type: none"> Physical health interventions (e.g., surgery)
Outcomes	<ul style="list-style-type: none"> Behavioral health-relevant outcomes include but are not limited to: mental health (e.g., depression, anxiety, trauma), substance use, and social and role functioning 	<ul style="list-style-type: none"> Physical health outcomes (e.g., blood pressure)
Setting	<ul style="list-style-type: none"> Behavioral health-friendly settings include but are not limited to: mental health treatment centers, medical care facilities in which behavioral health is integrated, criminal justice, education, social service 	<ul style="list-style-type: none"> NA
Measurement type	Include: <ul style="list-style-type: none"> Quantitative measures, typically self-report surveys, formulas, and equations 	<ul style="list-style-type: none"> Qualitative evaluation

Table 3. Mapping of scales and subscales to outer setting constructs.

Instrument	Scale or subscale mapped to construct				
	Outer setting domain	Cosmopolitanism	External policy and incentives	Patient needs and resources	Peer pressure
Nieboer and Stratling survey of structural characteristics ^a	Full scale	Full scale			Full scale
SHAY	Full scale		1 subscale		
EBP Survey	1 subscale				
Feinberg et al. measure of community ties	Full scale				
PSAT		1 subscale			
Luke et al. network relations measure		Full scale			
NCJTP Survey		2 subscales			
Smolders et al. cosmopolitanism measure		1 subscale			
Brookman-Frazee et al. survey		1 subscale			
SOCIS		1 subscale		4 subscales	
Measure of awareness of government initiatives			Full scale		
NSPO Survey			2 subscales		
CSP Director Survey			Full scale		
TCU-ORC				2 subscales	
Survey of organizational functioning				2 subscales	
Tool for measurement of ACT				1 subscale	
Total number of measures	4	7	4	4	1

Note. SHAY=State Health Authority Yardstick; EBP Survey=Evidence Based Practice Survey; PSAT=Program Sustainability Assessment Tool; NCJTP Survey=National Criminal Justice Treatment Practice Survey; SOCIS=Systems of Care Implementation Survey; NSPO=National Study of Physician Organizations survey; CSP Director Survey=Clinical Systems Project Director Survey; TCU-ORC=Texas Christian University Organizational Readiness for Change; ACT=Assertive Community Treatment.

^aNieboer and Stratling’s survey of structural characteristics combined a number of scales and validated items. From the source article, it was not clear if these were considered subscales for a full validated scale, or separate validated scales. As some measures had psychometric information available, they were treated as separate scales rather than subscales.

known-groups validity, structural validity, or responsiveness. See Figure 1 for the individual ratings for each psychometric criterion and total score for the three measures of outer setting with information available.

Cosmopolitanism. Seven measures of cosmopolitanism were identified. Five of these measures were suitable for rating and had some psychometric information available. There was information available on internal consistency

and norms for three measures, predictive validity for two measures, and information on known-groups validity for one measure. There was no psychometric information available on convergent validity, discriminant validity, concurrent validity, structural validity, and responsiveness.

For measures of cosmopolitanism with information available, the median rating for internal consistency was “3-good,” “2-adequate” for norms, and “-1-poor” for known-groups validity and predictive validity. The median

Table 4. Description of measure characteristics.

Characteristic	Outer setting		Cosmopolitanism		External policy and incentives		Patient needs and resources		Peer pressure	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Country of origin										
US	3	75	5	71	3	75	4	100	0	0
Other	1	25	2	29	1	25	0	0	1	100
Concept defined										
Yes	3	75	6	86	3	75	4	100	1	100
No	1	25	1	14	1	25	0	0	0	0
Used on one occasion										
Yes	4	100	5	71	3	75	1	25	1	100
No	0	0	2	29	1	25	3	75	0	0
Number of subscales										
1	0	0	1	14	0	0	1	25	0	0
2–5	0	0	1	14	1	25	3	75	0	0
6 or more	1	25	1	14	1	25	0	0	0	0
Not specified	3	75	4	57	2	50	0	0	1	100
Number of items										
1–5	0	0	0	0	0	0	1	25	1	100
6–10	0	0	0	0	0	0	0	0	0	0
11 or more	3	75	5	71	2	50	3	75	0	0
Not specified	1	25	2	29	2	50	0	0	0	0
Setting ^a										
State mental health	1	25	0	0	0	0	0	0	0	0
Inpatient psychiatry	0	0	0	0	0	0	1	25	0	0
Outpatient community	1	25	2	29	2	50	3	75	0	0
School mental health	0	0	1	14	0	0	2	50	0	0
Residential care	1	25	2	29	0	0	1	25	0	0
Other	3	75	7	100	3	75	4	100	1	100
Level ^a										
Consumer	0	0	0	0	0	0	0	0	0	0
Organization	0	0	0	0	0	0	0	0	0	0
Clinic/site	0	0	1	14	1	25	2	50	0	0
Provider	1	25	5	71	2	50	3	75	1	100
System	1	25	0	0	1	25	0	0	0	0
Team	0	0	0	0	0	0	1	25	0	0
Director	2	50	1	14	0	0	2	50	0	0
Supervisor	2	50	4	57	1	25	2	50	1	100
Other	0	0	2	29	0	0	1	25	0	0
Target problem ^a										
General mental health	3	75	3	43	3	75	2	50	1	100
Anxiety	0	0	1	14	0	0	1	25	0	0
Depression	0	0	1	14	0	0	1	25	0	0
Suicidal ideation	0	0	0	0	0	0	0	0	0	0
Alcohol use disorder	1	25	1	14	0	0	1	25	0	0
Substance use disorder	2	50	2	29	1	25	3	75	0	0
Behavioral disorder	1	25	0	0	0	0	1	25	0	0
Mania	0	0	0	0	0	0	0	0	0	0
Eating disorder	0	0	0	0	0	0	0	0	0	0
Grief	0	0	0	0	0	0	0	0	0	0
Tic disorder	0	0	0	0	0	0	0	0	0	0
Trauma	0	0	1	13	0	0	0	0	0	0
Other	0	0	2	29	1	25	1	25	0	0

(Continued)

Table 4. (Continued)

Characteristic	Outer setting		Cosmopolitanism		External policy and incentives		Patient needs and resources		Peer pressure	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
EPIS phase ^a										
Exploration	3	75	3	43	0	0	2	50	1	100
Preparation	1	25	1	14	1	25	1	25	0	0
Implementation	0	0	4	57	2	50	3	75	0	0
Sustainment	0	0	1	14	0	0	0	0	0	0
Outcome assessed ^a										
Acceptability	0	0	0	0	0	0	0	0	0	0
Appropriateness	0	0	0	0	0	0	0	0	0	0
Adoption	0	0	1	14	0	0	1	25	0	0
Cost	0	0	0	0	0	0	0	0	0	0
Feasibility	0	0	0	0	0	0	0	0	0	0
Fidelity	1	25	0	0	1	25	0	0	0	0
Penetration	0	0	0	0	0	0	0	0	0	0
Sustainability	0	0	0	0	0	0	0	0	0	0

Note. EPIS = Exploration, Adoption/Preparation, Implementation, Sustainment.

^aSome category levels are not mutually exclusive, for example, measures were used in multiple settings. This applies to the levels of analysis, populations, EPIS phases, and outcomes predicted.

Table 5. Psychometric information available on scales and subscales.

Psychometric property	Outer setting		Cosmopolitanism		External policy and incentives		Patient needs and resources		Peer pressure	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Unsuitable for rating	1	25	2	29	1	25	0	0	0	0
Measures with no information available	1	25	0	0	2	50	1	25	1	100
Measures with information available	2	50	5	71	1	25	3	75	0	0
Internal consistency	2	50	3	43	1	25	3	75	0	0
Convergent validity	0	0	0	0	0	0	1	25	0	0
Discriminant validity	0	0	0	0	0	0	0	0	0	0
Known-groups validity	0	0	1	14	0	0	1	25	0	0
Predictive validity	2	50	2	29	0	0	1	25	0	0
Concurrent validity	1	25	0	0	0	0	0	0	0	0
Structural validity	0	0	0	0	0	0	0	0	0	0
Responsiveness	0	0	0	0	0	0	0	0	0	0
Norms	2	50	3	43	1	25	2	50	0	0

rating of “-1-poor” for known-groups validity was based on a single subscale: the collaborative process subscale within Brookman-Frazer et al.’s (2016) web-based survey of the use of research community partnerships. Similarly, the median rating of “-1-poor” for predictive validity was based on median ratings of two subscales contained in the National Criminal Justice Treatment Practice (NCJTP) survey (Taxman et al., 2007). See Figure 2 for the individual ratings for each psychometric criterion with information available and total scores for each measure of cosmopolitanism.

Nieboer and Stratling’s communication scale had the highest overall psychometric rating score among measures

of cosmopolitanism used in behavioral health (psychometric total score = 5; maximum possible score = 36), with ratings of “3-good” for norms and “2-adequate” for predictive validity (Nieboer & Stratling, 2012). However, it is important to note that these scores are from a single study and there was no information available on any of the other psychometric criteria.

External policies and incentives. Four measures of external policies and incentives were identified in behavioral health research. Three measures were suitable for rating; however, only one measure had any psychometric information available. There was information available about internal

Table 6. Summary statistics for instrument ratings.

Psychometric property	Outer setting		Cosmopolitanism		External policy and incentives		Patient needs and resources		Peer pressure	
	Median	Range	Median	Range	Median	Range	Median	Range	Median	Range
Internal consistency	2	1,3	3	1,3	1	1	2	2,3	—	—
Convergent validity	—	—	—	—	—	—	2	2	—	—
Discriminant validity	—	—	—	—	—	—	—	—	—	—
Known-groups validity	—	—	-1	-1	—	—	3	3	—	—
Structural validity	—	—	—	—	—	—	—	—	—	—
Predictive validity	2	2,3	-1*	-1,2	—	—	-1	-1	—	—
Concurrent validity	1	1	—	—	—	—	—	—	—	—
Responsiveness	—	—	—	—	—	—	—	—	—	—
Norms	2	2	2	-1,3	4	4	2	2	—	—

Note. The results are based on those measures which had psychometric information available for rating: outer setting (*n* = 3 measures), cosmopolitanism (*n* = 5 measures), external policy and incentives (*n* = 1 measure), and patient needs and resources (*n* = 3 measures). Median calculated, excluding measures deemed unsuitable for rating or where psychometric information is not available (i.e., rated 0).

*Where the median of two scores would equal “0” or when rounded down would equal “0” (e.g., a score of -1 and 1, a score of 0.5), the lower score was taken.

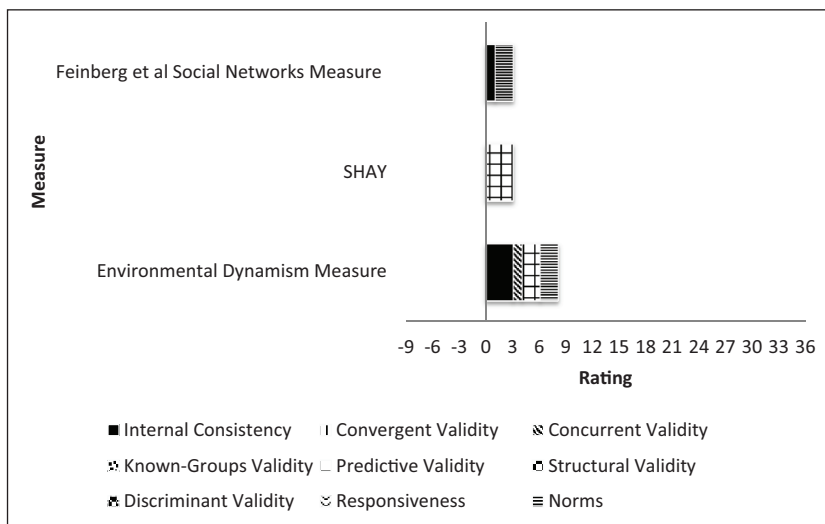


Figure 1. Ratings for each psychometric criterion and total score for measures of outer setting.

Note. This figure illustrates the measures (*n* = 3) and criteria for which there was psychometric information available for rating.

consistency and norms from only one measure, the National Survey of Physician Organizations (NSPO) survey (Ramsey et al., 2016). For this measure, the median rating was “4-excellent” for norms and “1-minimal” for internal consistency, based on scores from two relevant subscales: public reporting index and pay-for-performance index.

As the only measure with psychometric information, this measure had the highest overall psychometric rating score (total score=5; maximum possible score=36). However, the ratings for internal consistency varied between subscales: a rating of “2-adequate” for the public reporting index subscale and “1-minimal” for the pay-for-performance index subscale, the median of which was rounded down to a rating of “1-minimal” for the overall scale.

Patient needs and resources. Four measures of patient needs and resources were identified in behavioral health research. All four measures were suitable for rating but only three measures had some psychometric information available. There was information available on internal consistency for three measures, norms for two measures, and information on convergent validity, predictive validity, and known-groups validity for one measure. There was no psychometric information available on discriminant validity, concurrent validity, structural validity, and responsiveness from any of the measures.

For those measures of patient needs and resources with information available, the median rating was “3-good” for known-groups validity and “2-adequate” for internal

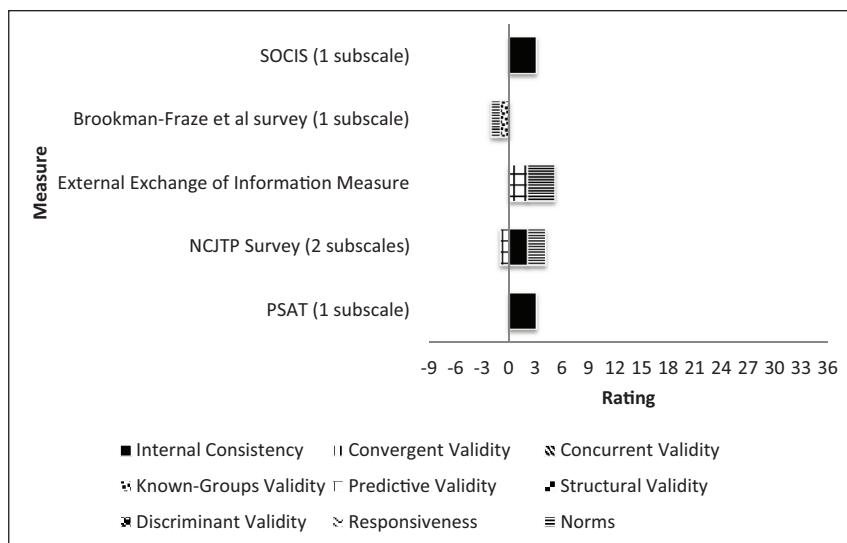


Figure 2. Ratings for each psychometric criterion and total score for measures of cosmopolitanism. Note. This figure illustrates the measures (n = 5) and criteria for which there was psychometric information available for rating.

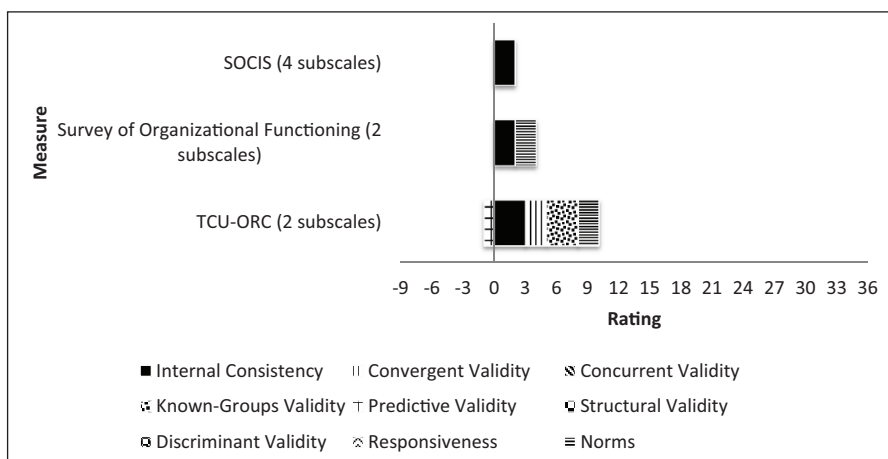


Figure 3. Ratings for each psychometric criterion and total score for measures of patient needs and resources. Note. This figure illustrates the measures (n = 3) and criteria for which there was psychometric information available for rating.

consistency, convergent validity, and norms. The median rating for predictive validity was “-1-poor” (Table 4). Note the median ratings of “3-good” for known-groups validity and “2-adequate” for convergent validity were based on two subscales within the Texas Christian University Organizational Readiness for Change (TCU-ORC) (Lehman et al., 2002). The median rating of “-1-poor” for predictive validity was also based on these two subscales. See Figure 3 for the individual ratings for each psychometric criterion and total scores for the three measures of patient needs and resources with information available.

Thus, one measure, the TCU-ORC, contributed most of the psychometric information on patient needs and resources and had the highest overall psychometric rating score (psychometric total score=9; maximum possible score=36). It had ratings of “3-good” for internal consistency and known

groups, “2-adequate” for convergent validity and norms, and “-1-poor” for predictive validity as previously mentioned.

Peer pressure. One measure of peer pressure was identified: Nieboer and Stratling’s measure of environmental competitiveness (Nieboer & Stratling, 2012). While it was suitable for rating, there was no psychometric information available across any of the nine criteria. As a result, there are no ratings of the quality of information available for this measure.

Discussion

The aim of this systematic review was to identify and rate the psychometric quality of measures used to assess outer setting constructs in behavioral and mental health research.

Overall, 20 measures were used in full or in part to assess the outer setting domain and/or its CFIR-delineated constructs: cosmopolitanism, external policies and incentives, patient needs and resources, and peer pressure. Most were distinct subscales located within full measures that assessed a range of implementation factors (e.g., TCU-ORC). The psychometric properties of identified measures were typically not reported. No single measure reported on more than five of the nine criteria. None of the measures had psychometric information available on discriminant validity, structural validity, or responsiveness. Similar to the findings of another review of implementation measures in public health contexts (Clinton-McHarg et al., 2016), few measures reported criterion-predictive validity or known-groups validity. Internal consistency and norms were the psychometric criterion most commonly assessed by measures of outer setting constructs used in behavioral health research. When measures did report psychometric data, it was typically rated as “minimal” or “adequate” across criteria. Only one criterion, norms, was rated as “excellent,” based on a single measure, the NSPO survey developed by Ramsey et al. (2016) which contained two scales to assess external policy and incentives. In this survey, measures of central tendency and distribution for the total score were based on a large ($n > 500$) nationally representative sample of medical practices.

Of all the implementation domains systematically reviewed as part of the SIRC Instrument Review Project, the outer setting was the focus of the smallest number of identified measures. This suggests that the outer setting has been neglected compared to levels that are easier to manipulate and measure. Given the dependence of behavioral and mental health care on outer setting factors (Purtle et al., 2017), consistent measurement that is theory-based and linked to implementation and client outcomes is of particular importance in this field. Our results align with the findings from other reviews which examined measures of implementation determinants at multiple levels in general health (Chaudoir et al., 2013), public and community health (Clinton-McHarg et al., 2016), and state mental health systems (Chor et al., 2015) where outer setting or “structural” level constructs were less frequently assessed. This is perhaps not surprising as policy is understudied in implementation research, at least in the United States where less than 10% of dissemination and implementation research funded by the National Institutes of Health (NIH) between 2007 and 2014 focussed on policy (Purtle et al., 2015). And yet, most measures of outer setting identified in this systematic review originated in the United States, suggesting other countries may face even greater challenges when quantitatively assessing outer setting constructs or policy implementation. The methodological challenges include the relative lack of opportunities for experimental manipulation, lack of operationalization of the relevant units of analysis for data collection, and large sample sizes required to test the effect of outer setting

constructs on implementation outcomes (Lewis, Proctor, Brownson, 2018).

Comparing the individual measures identified in the various reviews suggests there is very little overlap in measures of outer context used in different fields. This may be due, in part, to the different organizing frameworks and terms for outer setting used in other reviews such as “external system” (Chor et al., 2015), “structural level factors” (Chaudoir et al., 2013), and “environment” (Kaplan et al., 2010). However, each of the reviews refers to the CFIR framework as the foundation for, or comparable to, to their own organizing structure. The results more likely reflect the fact that most measures identified in our review were used once, in a single study. The results also suggest that in different implementation settings, some constructs are prioritized for measurement more than others. The recent systematic review of measures used in public health and community settings (Clinton-McHarg et al., 2016), which also used CFIR as the organizing framework, identified only two of the same measures of outer setting constructs as this review: the Systems of Care Implementation Survey (SOCIS; Boothroyd et al., 2011) which was developed for child mental health services, and scales from the NCJTP Survey which was developed for use in correctional settings (Taxman et al., 2007). Similar to our findings, few studies adequately reported or assessed the psychometric properties of measures, and those that did demonstrated poor psychometric quality suggesting few ready-to-use measures from other fields like public and community health. The lack of overlap limits our ability to learn about the psychometric performance of measures in other settings and whether they can be readily adapted to different contexts.

Within the outer setting domain, measures were not distributed equally across constructs. The construct for which most measures were identified was cosmopolitanism, defined in CFIR as “the degree to which an organization is networked with other external organisations” (Damschroder et al., 2009). This is in contrast to the review of measures used in public health and community settings which also mapped to CFIR constructs, where most measures of outer setting determinants assessed peer pressure (Clinton-McHarg et al., 2016). Similarly, a review of the health care QI literature identified the comparable construct of *competition* as the most commonly assessed “environment” factor (Kaplan et al., 2010). In our systematic review of behavioral health research, only one measure was identified for peer pressure; a three-item scale of environmental competitiveness with no psychometric information available at the time of review (Nieboer & Strating, 2012). Our findings may extend from a recognition that behavioral health often requires interaction (collaboration, coordination, funding arrangements) among multiple sectors (e.g., health, substance abuse, and, for children’s mental health, sectors such as education and child welfare). Indeed, previous research has found that density of

interorganizational relations is associated with greater EBP implementation success (Palinkas et al., 2014).

In this systematic review, 16 of the 20 identified measures of outer setting constructs were considered eligible for rating. The four measures deemed unsuitable for rating were open-ended surveys or semi-structured interview formats. In the literature, however, a variety and mix of measurement types are commonly used to assess outer setting constructs (Chor et al., 2015; Cook et al., 2012; Kaplan et al., 2010). A study to operationalize Greenhalgh et al.'s model of implementation used a combination of survey and interview questions supplemented with administrative data to assess "outer context" (Cook et al., 2012). Similarly, in a systematic review of predictors of EBP adoption, half of "external system" measures identified were derived from computations of frequency data, based on reviews of state documents, or open-ended surveys or interviews with key informants (Chor et al., 2015). In their systematic review of factors associated with QI success, Kaplan et al. highlighted the range of measurement types used to assess "environment" factors, highlighting one example of objective indices known as the Herfindahl Hirschman Index (HHI) which accounts for both the number and relative size of competing organizations to assess competition. Importantly, the results of the review suggested that the influence of factors, such as competition, varied depending on the type of measure used. For example, perceived competition was identified as a predictor of adoption; however, more objective competition indicators such as the HHI were not (Kaplan et al., 2010).

In implementation research, the lack of clarity on the appropriate method of measurement for the target level of analysis can lead to an overreliance on, and inappropriate use of use of self-report individual-level measures (Lewis, Proctor, Brownson 2018). For some outer setting constructs, it could be argued that the use of objective indices and administrative data may represent a better match between the level of measurement and the level of interpretation. For example, some previous policy studies in behavioral health have simply summed the number of supportive conditions to be in place (using individual dichotomous variables) to derive an index that served as an independent variable relevant to the supportiveness or hospitability of the policy context for some type of research-based practice, see, for example, Bruns et al. (2019). Certain constructs, such as external policies and incentives, do not reflect latent constructs and so are more appropriately measured objectively using administrative data which may also be more efficient and pragmatic (Cook et al., 2012). The appropriate measurement of other outer setting constructs is less clear cut. While there are objective measures of competition available such as the aforementioned HHI, there are important conceptual considerations such as identifying which organizations are considered competitors and so create mimetic pressure,

and practical considerations such as the availability of data to construct such indices (Baker, 2001). It could be argued there are objective measures of patients' need and resources such as indicators of disease prevalence or socioeconomic status. However, such measures may not accurately capture an organization's awareness of those needs and resources in line with the CFIR definition, and so surveys of staff and managements' perceptions may still be appropriate. For cosmopolitanism, social network analysis may offer a promising approach to capture the extent to which an organization is networked with others and assess how this influences implementation outcomes. In the main, these analyses are based on self-report surveys (Glegg et al., 2019). Acknowledging the practical considerations when collecting data, researchers should not assume "one size fits all" outer setting constructs. In order to advance the field, we recommend more careful consideration of the nature of individual constructs within the outer setting domain and the level of analysis in studies, to ensure appropriate selection of measures including self-report surveys.

Strengths and limitations

Previous reviews of implementation measures that included outer setting constructs have focused solely on a select number of psychometric properties (Chaudoir et al., 2013; Clinton-McHarg et al., 2016). The current systematic review rated the information available across nine psychometric criteria. Also, this systematic review identifies where constructs were assessed using full scales or individual subscales, giving a more accurate representation of the comprehensiveness of measures used to assess outer setting constructs. Most outer setting constructs were assessed using particular subscales within larger measures.

This is the first systematic review to focus exclusively on measures of outer setting constructs used in behavioral health research. The limited overlap with measures used in other fields such as public health highlights the potential to select and test relevant measures outside the behavioral health literature. While we are drawing these conclusions based on our review of measures used in mental and behavioral health, the similarity of findings across reviews suggests this approach is applicable to implementation research in other fields. We did not assess measures published in the gray literature; therefore, potentially relevant measures could have been overlooked. However, it is likely that the studies published in peer-reviewed journals represent the best available information on the psychometric quality of measures.

Another limitation of this 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 SIRC Instrument Review Project, it took the research

team nearly 2 years to screen articles, extract data, apply the rating system, and complete this manuscript. This systematic review is part of a larger project to identify measures of all implementation constructs associated with the 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. During the time between when we conducted our searches and when we finalized our data, it is possible that new measures of implementation outcomes were developed or the psychometric properties of included measures have been tested in more recent behavioral health research. 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.

Conclusion

To our knowledge, this is the first systematic review to concentrate solely on the availability and psychometric quality of measures used in behavioral and mental health research to assess outer setting constructs. This domain had the fewest measures identified across the SIRC Instrument Review Project and most measures found had little or no psychometric information available. Where information was available, psychometric properties were rated as minimal or adequate, indicating significant work for the field to ensure use of quality measures. Our results indicate that the outer context has been neglected compared to levels that are easier to manipulate and measure. To advance the field of implementation science toward a testable theory, there is a need to operationalize and measure all levels of implementation context. While outer setting factors may be difficult to influence, research has also found conditions in the outer setting that are both influential and mutable (Bruns et al., 2019; Palinkas et al., 2014). Thus, measuring these constructs is valuable to inform problem solving and to guide the selection of implementation strategies that are more likely to align with or leverage aspects of the outer setting.

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: C.C.L. is both an author of this manuscript and an editor of the journal, *Implementation Research and Practice*. Due to this conflict, C.C.L. was not involved in the editorial or review process for this manuscript.


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 C.C.L. as principal investigator. SMCh time on the manuscript was supported by the Fulbright Commission and Health Research Board Health Impact Scholarship [2017]. We would also like to acknowledge the numerous undergraduate research assistants (RAs) who contributed countless hours to the larger measures initiative by the Society for Implementation Research Collaboration (SIRC). Indiana University RAs listed in alphabetical order: Hayley Ciosek, Caitlin Dorsey, Dorina Feher, Sarah Fischer, Amanda Gray, Charlotte Hancock, Hilary Harris, Elise Hoover, Taylor Marshall, Elizabeth Parker, Paige Schultz, Monica Schuring, Theresa Thymoski, Lucia Walsh, Kaylee Will, Rebecca Zauel, Wannu Zhou, Anna Zimmerman, and Nelson Zounlome. University of Montana RAs (undergraduate and graduate) listed in alphabetical order: Kaitlyn Ahlers, Sarah Bigley, Melina Chapman, May Conley, Lindsay Crosby, Bridget Gibbons, Eleana Joyner, Samantha Moore, Julie Oldfield, Kinsey Owen, Amy Peterson, and Mark Turnipseed. University of North Carolina RAs: Emily Haines and Connor Kaine.

ORCID iDs

Sheena McHugh  <https://orcid.org/0000-0002-6595-0491>

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

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

References

- 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>
- Baker, L. C. (2001). Measuring competition in health care markets. *Health Services Research*, 36(1 Pt. 2), 223–251. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089203/pdf/hsresearch00002-0059.pdf>
- Boothroyd, R. A., Greenbaum, P. E., Wang, W., Kutash, K., & Friedman, R. M. (2011). Development of a measure to assess the implementation of children’s systems of care: The systems of care implementation survey (SOCIS). *The Journal of Behavioral Health Services & Research*, 38(3), 288–302. <https://doi.org/10.1007/s11414-011-9239-x>
- Brookman-Frazee, L., Stahmer, A., Stadnick, N., Chlebowski, C., Herschell, A., & Garland, A. F. (2016). Characterizing the use of research-community partnerships in studies of evidence-based interventions in children’s community services. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(1), 93–104. <https://doi.org/10.1007/s10488-014-0622-9>
- Bruns, E. J., Parker, E. M., Hensley, S., Pullmann, M. D., Benjamin, P. H., Lyon, A. R., & Hoagwood, K. E. (2019). The role of the outer setting in implementation: Associations between state demographic, fiscal, and policy factors and use of evidence-based treatments in mental healthcare. *Implementation Science*, 14(1), Article 96. <https://doi.org/10.1186/s13012-019-0944-9>
- Chaudoir, S. R., Dugan, A. G., & Barr, C. H. (2013). Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures. *Implementation Science*, 8(1), Article 22. <https://doi.org/10.1186/1748-5908-8-22>

- Chor, K. H. B., Wisdom, J. P., Olin, S.-C. S., Hoagwood, K. E., & Horwitz, S. M. (2015). Measures for predictors of innovation adoption. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 545–573. <https://doi.org/10.1007/s10488-014-0551-7>
- Clinton-McHarg, T., Yoong, S. L., Tzelepis, F., Regan, T., Fielding, A., Skelton, E., Kingsland, M., Ooi, J. Y., & Wolfenden, L. (2016). Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the consolidated framework for implementation research: A systematic review. *Implementation Science*, 11(1), Article 148. <https://doi.org/10.1186/s13012-016-0512-5>
- Cook, J. M., O'Donnell, C., Dinnen, S., Coyne, J. C., Ruzek, J. I., & Schnurr, P. P. (2012). Measurement of a model of implementation for health care: Toward a testable theory. *Implementation Science*, 7, Article 59. <https://doi.org/10.1186/1748-5908-7-59>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(1), Article 50. <https://doi.org/10.1186/1748-5908-4-50>
- Davidoff, F., Batalden, P., Ogrinc, G., & Mooney, S. (2008). Publication guidelines for quality improvement studies in healthcare: Evolution of the SQUIRE Project. *Canadian Journal of Diabetes*, 32(4), 281–289. [https://doi.org/10.1016/S1499-2671\(08\)24008-6](https://doi.org/10.1016/S1499-2671(08)24008-6)
- Emmons, K. M., Weiner, B., Fernandez, M. E., & Tu, S.-P. (2012). Systems antecedents for dissemination and implementation: A review and analysis of measures. *Health Education & Behavior*, 39(1), 87–105. <https://doi.org/10.1177/1090198111409748>
- Feldstein, A. C., & Glasgow, R. E. (2008). A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *The Joint Commission Journal on Quality and Patient Safety*, 34(4), 228–243. [https://doi.org/10.1016/s1553-7250\(08\)34030-6](https://doi.org/10.1016/s1553-7250(08)34030-6)
- French, B., Thomas, L. H., Baker, P., Burton, C. R., Pennington, L., & Roddam, H. (2009). What can management theories offer evidence-based practice? A comparative analysis of measurement tools for organisational context. *Implementation Science*, 4(1), Article 28. <https://doi.org/10.1186/1748-5908-4-28>
- Glegg, S. M. N., Jenkins, E., & Kothari, A. (2019). How the study of networks informs knowledge translation and implementation: A scoping review. *Implementation Science*, 14, Article 34. <https://doi.org/10.1186/s13012-019-0879-1>
- Kaplan, H. C., Brady, P. W., Dritz, M. C., Hooper, D. K., Linam, W. M., Froehle, C. M., & Margolis, P. (2010). The influence of context on quality improvement success in health care: A systematic review of the literature. *Milbank Quarterly*, 88(4), 500–559. <https://doi.org/10.1111/j.1468-0009.2010.00611.x>
- Lehman, W. E., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse and 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(1), 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., Stanick Halko, H., & Powell, B. J. (2018). An updated protocol for a systematic review of implementation-related measures. *Systematic Reviews*, 7, Article 66. <https://doi.org/10.1186/s13643-018-0728-3>
- Lewis, C. C., Proctor, E., & Brownson, R. C. (2018). Measurement issues in dissemination and implementation research. In R. C. Brownson, G. A. Colditz & E. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (2nd ed., pp. 229–244). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199751877.001.0001>
- Nieboer, A. P., & Strating, M. M. (2012). Innovative culture in long-term care settings: The influence of organizational characteristics. *Health Care Management Review*, 37(2), 165–174. <https://doi.org/10.1097/HMR.0b013e318222416b>
- Palinkas, L. A., Fuentes, D., Finno, M., Garcia, A. R., Holloway, I. W., & Chamberlain, P. (2014). Inter-organizational collaboration in the implementation of evidence-based practices among public agencies serving abused and neglected youth. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(1), 74–85. <https://doi.org/10.1007/s10488-012-0437-5>
- Purtle, J., Brownson, R., & Proctor, E. (2017). Infusing science into politics and policy: The importance of legislators as an audience in mental health dissemination research. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(2), 160–163. <https://doi.org/10.1007/s10488-016-0752-3>
- Purtle, J., Peters, R., & Brownson, R. C. (2015). A review of policy dissemination and implementation research funded by the National Institutes of Health, 2007–2014. *Implementation Science*, 11(1), Article 1. <https://doi.org/10.1186/s13012-015-0367-1>
- Ramsay, P. P., Shortell, S. M., Casalino, L. P., Rodriguez, H. P., & Rittenhouse, D. R. (2016). A longitudinal study of medical practices' treatment of patients who use tobacco. *American Journal of Preventive Medicine*, 50(3), 328–335. <https://doi.org/10.1016/j.amepre.2015.07.005>
- Smith, D. M., Damschroder, L. J., Kim, S. Y., & Ubel, P. A. (2012). What's it worth? Public willingness to pay to avoid mental illnesses compared with general medical illnesses. *Psychiatric Services*, 63(4), 319–324. <https://doi.org/10.1176/appi.ps.2010.00.036>
- Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. C. (2012). Bridging research and practice: Models for dissemination and implementation research. *American Journal of Preventive Medicine*, 43(3), 337–350. <https://doi.org/10.1016/j.amepre.2012.05.024>
- Taxman, F. S., Young, D. W., Wiersema, B., Rhodes, A., & Mitchell, S. (2007). The National Criminal Justice Treatment Practices survey: Multilevel survey methods and procedures. *Journal of Substance Abuse and Treatment*, 32(3), 225–238. <https://doi.org/10.1016/j.jsat.2007.01.002>
- Weiner, B. J., Amick, H., & Lee, S.-Y. D. (2008). Conceptualization and measurement of organizational readiness for change: A review of the literature in health services research and other fields. *Medical Care Research and Review*, 65(4), 379–436. <https://doi.org/10.1177/1077558708317802>

Appendix I

Construct PRISMA flowcharts

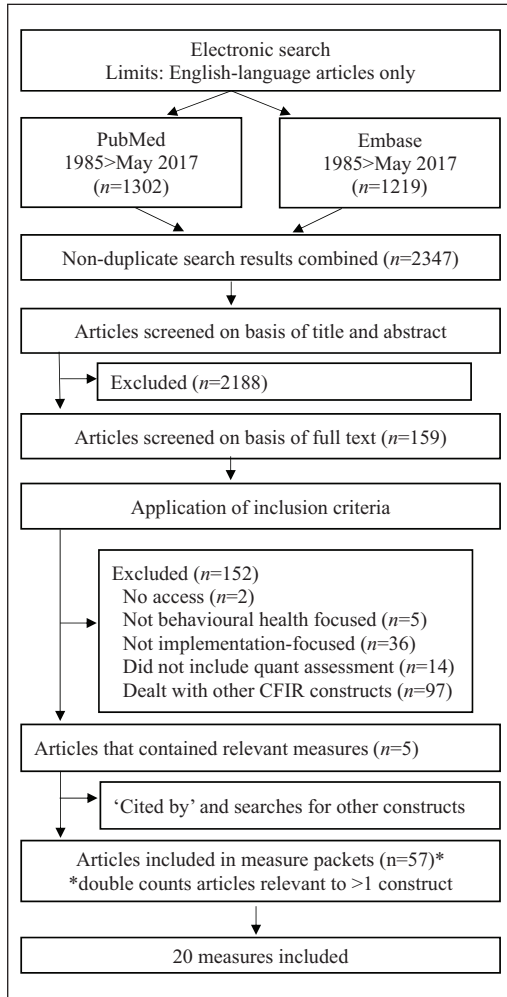


Figure A1. PRISMA enhanced systematic review flowchart for all constructs.

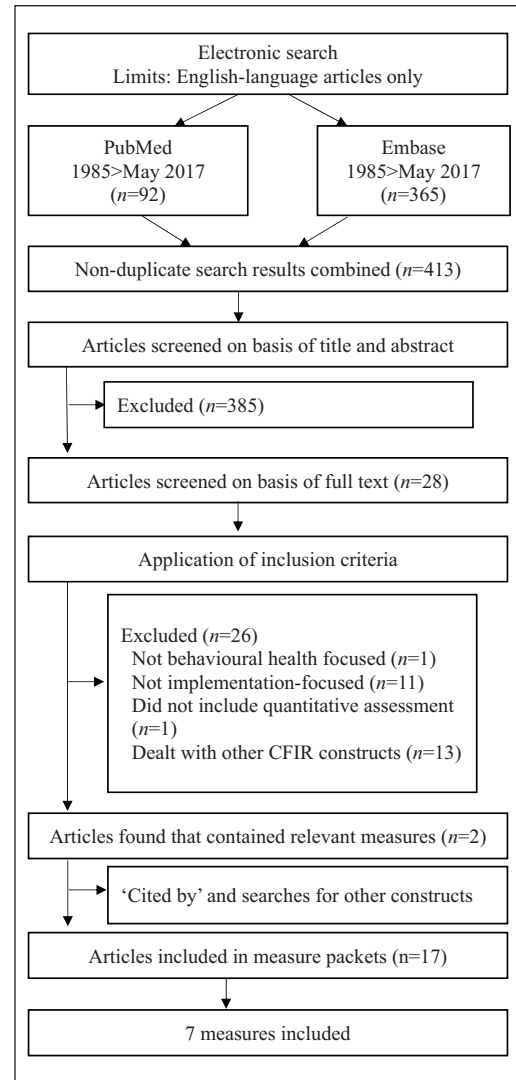


Figure A2. Cosmopolitanism PRISMA systematic review flowchart.

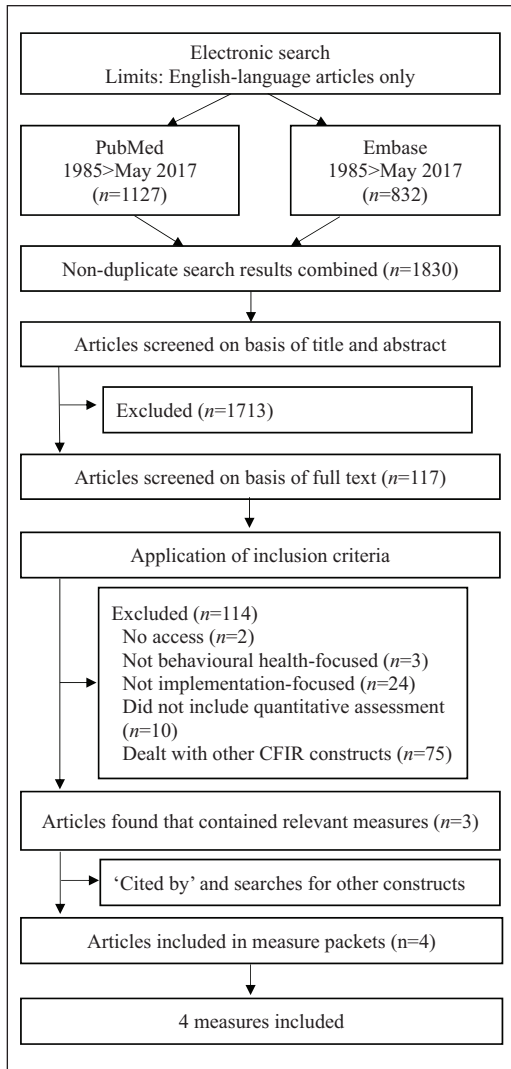


Figure A3. External policies and incentives PRISMA systematic review flowchart.

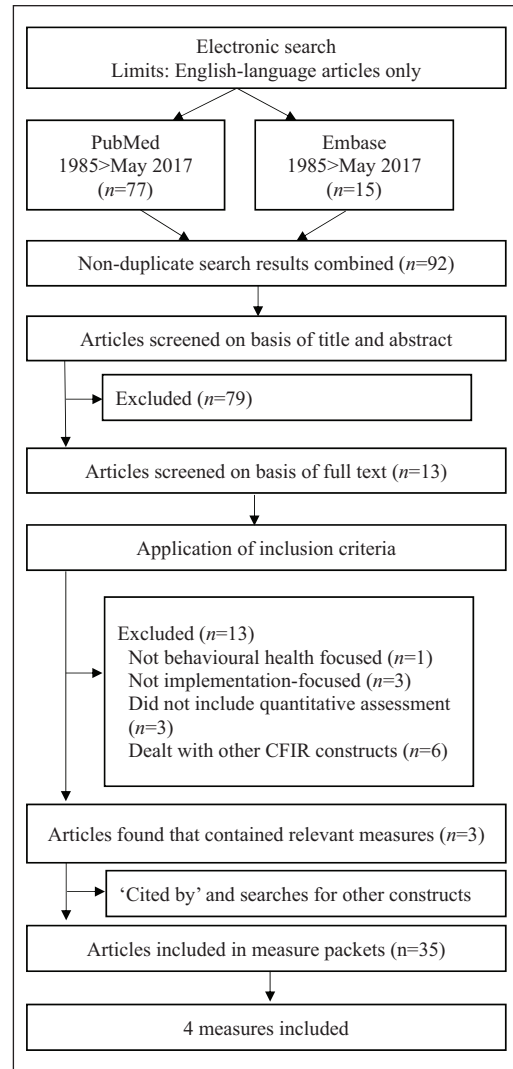


Figure A4. Patient needs and resources PRISMA systematic review flowchart.

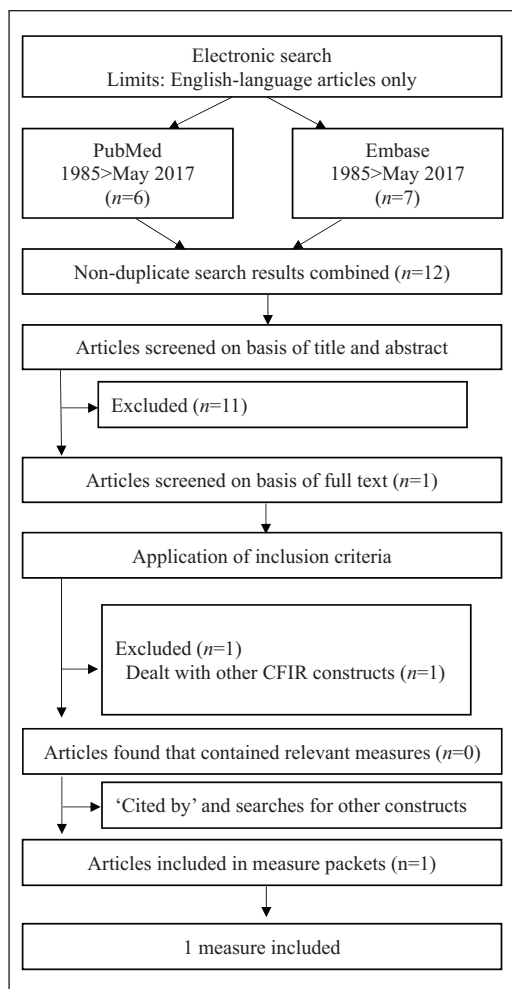


Figure A5. Peer pressure PRISMA systematic review flowchart.