



Measures of organizational culture, organizational climate, and implementation climate in behavioral health: A systematic review

Implementation Research and Practice
Volume 2: Jan-Dec 2021 1–29
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26334895211018862
journals.sagepub.com/home/irp

Byron J Powell¹, Kayne D Mettert², Caitlin N Dorsey²,
Bryan J Weiner³, Cameo F Stanick⁴, Rebecca Lengnick-Hall¹,
Mark G Ehrhart⁵, Gregory A Aarons⁶, Melanie A Barwick⁷,
Laura J Damschroder⁸ and Cara C Lewis²

Abstract

Background: Organizational culture, organizational climate, and implementation climate are key organizational constructs that influence the implementation of evidence-based practices. However, there has been little systematic investigation of the availability of psychometrically strong measures that can be used to assess these constructs in behavioral health. This systematic review identified and assessed the psychometric properties of measures of organizational culture, organizational climate, implementation climate, and related subconstructs as defined by the Consolidated Framework for Implementation Research (CFIR) and Ehrhart and colleagues.

Methods: Data collection involved search string generation, title and abstract screening, full-text review, construct assignment, and citation searches for all known empirical uses. Data relevant to nine psychometric criteria from the Psychometric and Pragmatic Evidence Rating Scale (PAPERS) were extracted: internal consistency, convergent validity, discriminant validity, known-groups validity, predictive validity, concurrent validity, structural validity, responsiveness, and norms. Extracted data for each criterion were rated on a scale from -1 (“poor”) to 4 (“excellent”), and each measure was assigned a total score (highest possible score = 36) that formed the basis for head-to-head comparisons of measures for each focal construct.

Results: We identified full measures or relevant subscales of broader measures for organizational culture ($n=21$), organizational climate ($n=36$), implementation climate ($n=2$), tension for change ($n=2$), compatibility ($n=6$), relative priority ($n=2$), organizational incentives and rewards ($n=3$), goals and feedback ($n=3$), and learning climate ($n=2$). Psychometric evidence was most frequently available for internal consistency and norms. Information about other

¹Brown School and School of Medicine, Washington University in St. Louis, St. Louis, MO, USA

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

³Department of Global Health and Department of Health Services, University of Washington, Seattle, WA, USA

⁴Hathaway-Sycamores Child and Family Services, Pasadena, CA, USA

⁵Department of Psychology, University of Central Florida, Orlando, FL, USA

⁶Department of Psychiatry, School of Medicine, University of California San Diego, La Jolla, CA, USA

⁷Hospital for Sick Children, Toronto, ON, Canada

⁸Ann Arbor VA Center for Clinical Management Research, Ann Arbor, MI, USA

Corresponding authors:

Byron J Powell, Brown School and School of Medicine, Washington University in St. Louis, One Brookings Drive, Campus Box 1196, St. Louis, MO 63130, USA.

Email: bjpowell@wustl.edu

Cara C Lewis, Kaiser Permanente Washington Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA 98101, USA.

Email: cara.c.lewis@kp.org



psychometric properties was less available. Median ratings for psychometric properties across categories of measures ranged from “poor” to “good.” There was limited evidence of responsiveness or predictive validity.

Conclusion: While several promising measures were identified, the overall state of measurement related to these constructs is poor. To enhance understanding of how these constructs influence implementation research and practice, measures that are sensitive to change and predictive of key implementation and clinical outcomes are required. There is a need for further testing of the most promising measures, and ample opportunity to develop additional psychometrically strong measures of these important constructs.

Plain Language Summary Organizational culture, organizational climate, and implementation climate can play a critical role in facilitating or impeding the successful implementation and sustainment of evidence-based practices. Advancing our understanding of how these contextual factors independently or collectively influence implementation and clinical outcomes requires measures that are reliable and valid. Previous systematic reviews identified measures of organizational factors that influence implementation, but none focused explicitly on behavioral health; focused solely on organizational culture, organizational climate, and implementation climate; or assessed the evidence base of all known uses of a measure within a given area, such as behavioral health–focused implementation efforts. The purpose of this study was to identify and assess the psychometric properties of measures of organizational culture, organizational climate, implementation climate, and related subconstructs that have been used in behavioral health-focused implementation research. We identified 21 measures of organizational culture, 36 measures of organizational climate, 2 measures of implementation climate, 2 measures of tension for change, 6 measures of compatibility, 2 measures of relative priority, 3 measures of organizational incentives and rewards, 3 measures of goals and feedback, and 2 measures of learning climate. Some promising measures were identified; however, the overall state of measurement across these constructs is poor. This review highlights specific areas for improvement and suggests the need to rigorously evaluate existing measures and develop new measures.

Keywords

Implementation science, measurement, reliability, validity, psychometric properties, organizational culture, organizational climate, implementation climate, mental health, behavioral health

Introduction

Because most behavioral health services are delivered within or through organizations (Aarons et al., 2018), organizational context plays a critical role in determining successful implementation of evidence-based practices (EBPs; Aarons et al., 2014, 2018; Glisson & Williams, 2015). Consequently, organizational context is included in ~95% of implementation frameworks (Tabak et al., 2012). The purpose of this systematic review was to identify and assess the psychometric properties of measures of organizational culture, organizational climate, and implementation climate used in behavioral health-related implementation studies. We drew upon conceptualizations of these constructs and related subconstructs from the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2009) and from the book by Ehrhart, Schneider, et al. (2014) (see Table 1).

Organizational culture

Organizational culture is defined as “. . . the shared values and basic assumptions that explain why organizations do what they do and focus on what they focus on” (Schneider et al., 2017, p. 468). There are debates regarding its definition and measurement (Aarons et al., 2018; Ehrhart, Schneider, et al., 2014; Kimberly & Cook, 2008; Schein,

2017; Verbeke et al., 1998). Verbeke and colleagues (1998) identified 54 definitions of organizational culture, and Aarons and colleagues (2018) described six different measures that use 3–10 different dimensions each to measure culture (e.g., involvement, adaptability, mission, attention to detail, aggressiveness, innovation, supportiveness, leadership, planning, communication, hierarchy, proficiency, and apathy), with very little overlap. In behavioral health, organizational culture has been empirically linked to attitudes toward EBPs, sustainment, access to services, service quality, staff turnover, and mental health outcomes (Glisson & Williams, 2015).

Organizational climate

Organizational climate is defined as “the shared meaning organizational members attach to the events, policies, practices, and procedures they experience and the behaviors they see being rewarded, supported, and expected” (Ehrhart, Schneider, et al., 2014, p. 69). Scholars differentiate molar organizational climate from focused climates. Molar conceptualizations refer to the extent to which leadership provides positive experiences for employees (Aarons et al., 2018) and can include dimensions such as engagement, functionality, and stress (Glisson et al., 2008). In behavioral health, molar organizational climate has been empirically linked to service quality, treatment planning decisions, attitudes toward EBPs,

staff turnover, and mental health outcomes (Glisson & Williams, 2015). Others have conceptualized and measured focused aspects of climate, such as an outcome (e.g., climate for safety) or organizational process (e.g., ethics, fairness) (Aarons et al., 2018).

Implementation climate

Implementation climate is a type of focused organizational climate defined as a summary of employees' shared perceptions of the extent to which their use of an innovation is rewarded, supported, and expected (Klein & Sorra, 1996). Strong implementation climates encourage use of EBPs by (1) ensuring employees are skilled in their use, (2) incentivizing the use of EBP and eliminating disincentives, and (3) removing barriers to EBP use (Klein & Sorra, 1996). Implementation climate differs from molar organizational climate in that it is innovation-specific and focuses on organizational members who are expected to use or directly support an innovation (Weiner et al., 2011). Implementation climate may be critical to improving EBP implementation (Williams et al., 2018, 2020). Williams and colleagues' (2020) 5-year panel analysis showed that organizations that improved from low to high levels of implementation climate had significantly greater increases in their clinicians' average EBP use.

Previous reviews

Reviews of the organizational culture and organizational climate measures vary along several dimensions. First, some have systematically reviewed the literature (e.g., Allen et al., 2017; Chaudoir et al., 2013; Clinton-McHarg et al., 2016), whereas others have selectively reviewed measures at the discretion of the authors (e.g., Glisson & Williams, 2015; Kimberly & Cook, 2008; Schneider et al., 2013). Second, they range from narrow (e.g., organizational culture only; Jung et al., 2009; King & Byers, 2007; Scott et al., 2003) to broad assessments (e.g., a list of organizational characteristics; Allen et al., 2017; Brennan et al., 2012; Chaudoir et al., 2013). Third, they may or may not report psychometric properties and/or report them with varying degrees of granularity (e.g., Schneider et al., 2013, and Kimberly & Cook, 2008, do not report psychometric properties; Chaudoir et al., 2013, focused on criterion validity). Finally, they may (e.g., Clinton-McHarg et al., 2016) or may not (e.g., Gershon et al., 2004; Jung et al., 2009) be informed by a conceptual framework.

Two published systematic reviews examined psychometric properties of measures for constructs within the "inner setting" domain of the CFIR (Damschroder et al., 2009). Clinton-McHarg and colleagues (2016) examined quantitative measures developed for public health and community settings and located 51 measures. Most did not report on psychometric properties and those that did typically fell below accepted standards (Clinton-McHarg

et al., 2016). Allen et al. (2017) identified 83 measures of the inner setting and the two constructs with the most measures were readiness for implementation and organizational climate. However, only 46% of studies ($n=35$) included information about psychometric properties, and of those, 94% (33/35) described reliability and 71% (25/35) reported validity.

Aims and contribution of the current study

The current study sought to identify and assess the psychometric properties of measures of organizational culture, organizational climate, and implementation climate used in behavioral health-related implementation studies. This review contributes to the implementation and behavioral health literatures by (1) focusing explicitly on the assessment of these constructs within behavioral health; (2) identifying measures for key constructs and subconstructs of the widely used CFIR (Damschroder et al., 2009); and (3) rigorously assessing evidence of measures' psychometric strength by using the Psychometric and Pragmatic Evidence Rating Scale (PAPERS; Lewis, Mettert, et al., 2018; Stanick et al., 2021). Our intent is to inform researchers', EBP purveyors' (Proctor et al., 2019), implementation support practitioners' (Albers et al., 2020), and other stakeholders' selection of high-quality measures, and to highlight areas in which further development and testing of measures is necessary.

Methods

Design overview

Data for this systematic review come from a project funded by the U.S. National Institute of Mental Health, which included multiple systematic reviews that identified implementation determinant (Damschroder et al., 2009) and outcome (Proctor et al., 2011) measures that were used within implementation studies in behavioral health (Lewis et al., 2018). The protocol for that study has been published elsewhere (Lewis et al., 2018). This systematic review was conducted in three phases. Phase I, data collection, included five steps: (1) search string generation, (2) title and abstract screening, (3) full text review, (4) construct assignment, and (5) measure-forward searches. Phase II, data extraction, consisted of coding relevant psychometric data, and Phase III involved data analysis.

Phase I: data collection

Literature searches were conducted in PubMed and Embase using search strings curated in consultation from PubMed support specialists and a library scientist. PubMed and Embase are commonly recommended for systematic reviews in health (Bramer et al., 2017; Higgins et al.,

2021). Other databases such as PsycINFO were considered, but pilot testing revealed that search yields were low and did not identify a substantial number of studies for the constructs of interest. Search terms focused on: (1) implementation; (2) measurement; (3) EBP; (4) behavioral health; and (5) organizational culture and implementation climate. Our conceptualization of organizational culture and implementation climate was guided by the CFIR (Damschroder et al., 2009), and included search terms for organizational culture, implementation climate, and related subconstructs: tension for change, compatibility, relative priority, organizational incentives and rewards, goals and feedback, and learning climate (see Table 1). Table 2 includes a complete listing of search terms for PubMed and Embase. Articles published in English from 1985 onwards were included in the search. Searches were completed from 3 April 2017 to 25 May 2017.

Identified titles and abstracts were screened, followed by full-text review to confirm relevance to study parameters. We included empirical studies that contained one or more quantitative measures of the target constructs if they were used in an evaluation of an implementation effort in a behavioral health context. See Table 3 for inclusion/exclusion criteria, and Appendix 1 for PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowcharts (Figures 10 to 17).

The next step involved construct assignment, in which trained research specialists mapped measures and/or their subscales to the target constructs based on the authors' conceptualization of the measure and content expert coding. Inherent to our search approach, measures of the target constructs could be identified through systematic reviews of related constructs (e.g., organizational readiness for change) conducted in the parent study (Lewis et al., 2018). For example, a subscale from Organizational Readiness for Change Assessment ("Staff Culture") by Helfrich et al. (2009) was identified in a review of measures of organizational readiness for change (Weiner et al., 2020). The CFIR does not include molar organizational climate; thus, we did not conduct a search specifically for that construct. However, our searches for organizational culture and implementation climate identified a number of measures of molar organizational climate. We attempted to maintain a conceptual distinction between measures of molar organizational climate or focused organizational climates (e.g., risk taking climate; Cook et al., 2012) and the more intervention-specific implementation climate construct as described by Weiner et al. (2011). Thus, we ultimately categorized measures into nine different constructs: organizational culture, organizational climate, implementation climate, tension for change, compatibility, relative priority, organizational incentives and rewards, goals and feedback, and learning climate (Table 1).

Finally, "measure-forward" searches were conducted in May 2019 for each measure to identify empirical articles that used the measure in behavioral health implementation research. These searches were conducted using the "cited-by" feature in PubMed and Embase and by searching for measures' formal names as available.

Phase II: data extraction

Next, articles were compiled into "measure packets," including the measure itself (as available), the measure development article (or article with the first empirical use in a behavioral health context), and all identified empirical uses of the measure in behavioral health-related implementation efforts. Trained-research specialists reviewed each article and electronically extracted information relevant to nine psychometric rating criteria from the PAPERS (Lewis, Mettert, et al., 2018; Stanick et al., 2021): (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 (Table 4). Data were collected on both full measure and subscale levels. If a full measure was relevant to a target construct, we reported psychometric evidence for the full measure. However, if only subscales of a broader measure were relevant, we reported psychometric evidence at the subscale level. We use the term "measures" throughout this article to refer to both full measures and subscales; however, the distinction between the two is maintained by using formal names of measures and subscales in relevant tables and figures.

After PAPERS relevant data were extracted (Lewis, Mettert, et al., 2018; Stanick et al., 2021), each criterion was rated using the following scale for which nuanced anchors established: "poor" (-1), "none" (0), "minimal/emerging" (1), "adequate" (2), "good" (3), or "excellent" (4). Ratings were summarized using a "rolled up median" approach in an effort to assign a single score for each criterion. This is more reflective of the range of measure performance than often used "top score" or "worst score counts" methods (Lewis et al., 2015; Terwee et al., 2012). If a measure was unidimensional or the measure had only one rating for a criterion, then this value was the final rating. If a measure had multiple ratings for a criterion across several articles, we calculated the median score to generate the final rating. For example, if a measure was used in five different studies, each of which included evidence of internal consistency, we calculated the median to determine that measures' final rating of internal consistency. 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 a 2). In cases where the median of two scores would equal "0" (e.g., a score of -1 and 1), the lower would be taken (e.g., -1). This approach results in a conservative rating.

In addition to assessing psychometric properties, we extracted: (1) whether the measure was used more than once, (2) country of origin, (3) setting (e.g., inpatient psychiatry, outpatient), (4) level of analysis (e.g., consumer, organization, provider), (5) population (e.g., general mental health, anxiety, depression), and (6) stage of implementation as defined by the exploration, adoption/preparation, implementation, sustainment model (Aarons et al., 2011).

Phase III: data analysis

Simple statistics (frequencies, medians, ranges) were calculated to report on the presence and quality of psychometric data. Each measure was assigned a total score based upon the nine PAPERS criteria (highest possible score = 36). Bar charts were generated to display head-to-head comparisons across all measures within a given construct.

Results

Overview

Table 5 provides descriptive information. Table 6 shows availability of psychometric evidence. Table 7 includes the median and range of ratings of psychometric properties for measures with psychometric information available (i.e., those with non-zero ratings on the PAPERS criteria; Lewis, Mettert, et al., 2018; Stanick et al., 2021). Individual ratings for all measures are detailed in Table 8 and in head-to-head bar graphs in Figures 1 to 9.

Organizational culture

We identified 21 measures of organizational culture, 18 of which are subscales of broader measures (e.g., Organizational Readiness for Change Assessment–Staff Culture Scale; Helfrich et al., 2009). Measures were primarily developed in the United States (95%); used more than once (76%); used most frequently in outpatient community mental health (86%) and residential care settings (71%); administered most frequently at the provider (100%) or supervisor (52%) levels; used within general mental health (86%), alcohol use (57%), or substance use disorder (67%) services; and were used most frequently at the exploration (71%) and implementation (43%) phases.

Evidence of internal consistency was available for 18 measures, convergent validity for two measures, known-groups validity for two measures, predictive validity for 12 measures, concurrent validity for two measures, structural validity for one measure, responsiveness for one measure, and norms for 20 measures. No psychometric evidence was available for discriminant validity.

The median rating for internal consistency was “2—adequate,” for convergent validity “2—adequate,” for known-groups validity “–1—poor,” for predictive validity “1—minimal,” for concurrent validity “–1—poor,” for structural validity “2—adequate,” and for norms “2—adequate.” The median rating of “2—adequate” for structural validity was based on the rating of just one measure: the Organizational Social Context–Culture Scale (Glisson et al., 2008).

Table 1. Construct definitions.

Construct	Definition
Organizational culture	“Norms, values, and basic assumptions of a given organization”
Organizational climate	“The shared meaning organizational members attach to the events, policies, practices, and procedures they experience and the behaviors they see being rewarded, supported, and expected” (Ehrhart, Schneider, et al., 2014, p. 69)
Implementation climate	“The absorptive capacity for change, shared receptivity of involved individuals to an intervention and the extent to which use of that intervention will be rewarded, supported, and expected within their organization”
Tension for change	“The degree to which stakeholders perceive the current situation as intolerable or needing change”
Compatibility	“The degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals’ own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems”
Relative priority	Individuals’ shared perception of the importance of the implementation within the organization”
Organizational incentives and rewards	“Extrinsic incentives such as goal-sharing awards, performance reviews, promotions, and raises in salary and less tangible incentives such as increased stature or respect”
Goals and feedback	“The degree to which goals are clearly communicated, acted upon, and fed back to staff and alignment of that feedback with goals”
Learning climate	“A climate in which: a) leaders express their own fallibility and need for team members’ assistance and input; b) team members feel that they are essential, valued, and knowledgeable partners in the change process; c) individuals feel psychologically safe to try new methods; and d) there is sufficient time and space for reflective thinking and evaluation”

Note. Definitions from all constructs aside from organizational climate are drawn from Additional File 3 of Damschroder et al. (2009). The definition of organizational climate is taken from Ehrhart, Schneider, et al. (2014).

Table 2. Database search terms.

PubMed	
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 guidelined[tiab] OR guidelinedevelopment[tiab] OR guidelinei[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 guidelines[tiab] OR guidelines that[tiab] OR guidelines trade[tiab] OR guidelines wiki[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 surveysselect[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 questionnaireeqp[tw] OR questionnaireforpediatric[tw] OR questionnairegrtr[tw] OR questionnairehas[tw] OR questionnaireitaq[tw] OR questionnairel02[tw] OR questionnairemcesqscale[tw] OR questionnairenurse[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 questionnairesdq l adolescent[tw] OR questionnairess[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’s[tiab] OR measurement l [tiab] OR measuremental[tiab] OR measurementd[tiab] OR measurementend[tiab] OR measurementexhaled[tiab] OR measurementf[tiab] OR measurementin[tiab] OR measurementis[tiab] OR measurementis[tiab] OR measurementkomputation[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 measurementts[tiab] OR measurementusing[tiab] OR measurementws[tiab]) OR measures[tiab] OR inventory[tiab]) AND

(Continued)

Table 2. (Continued)

PubMed	
Search term	Search string
Behavioral 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 psychiatristes[tw] OR psychiatristis[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 "English"[Language] AND 1985[PDAT]: 3000[PDA T] AND
Culture	(implementation[tw] AND culture[tw]) OR "organizational implementation" OR "organizational norms" OR "organizational value*" OR "organizational assumptions" OR "organizational implementation"
Implementation climate	"implementation climate"[tw] OR "psychological climate"[tw] OR ("absorptive capacity" AND change[tw]) OR receptivity[tw]
Tension for change	"tension for change" OR "need for change" OR "climate for change"
Compatibility	"compatibility of climate" OR "organizational fit" OR "organizational fitness"
Relative priority	"relative priority" OR ("shared perception*" [tw] AND importance[tw])
Organizational incentives and rewards	"organizational incentives" OR "organizational rewards" OR "extrinsic incentives" OR "goal-sharing awards" OR "performance review*" [tw] OR promotion[tw] OR "goal sharing"[tw]
Goals and feedback	(goals[tw] OR "goals and objectives"[tw] AND feedback[tw]) OR "audit and feedback" OR "supervisor feedback" OR "goal* feedback*" OR "clinical evaluation" OR "clinical feedback" OR "performance feedback" OR "performance evaluation" OR "economic evaluation"
Learning climate	"learning climate" OR "learning capability" OR "peer collaboration"
Embase	
Search term	Search string
Organizational culture	(implementation AND culture) OR 'organizational implementation' OR 'organizational norms' OR 'organizational value*' OR 'organizational assumptions' OR 'organizational implementation'
Implementation climate	'implementation climate' OR 'psychological climate' OR ('absorptive capacity' AND change) OR receptivity OR 'organizational climate' OR 'organization climate' OR 'work environment' OR 'work attitudes' OR depersonalization OR 'emotional exhaustion' OR 'role conflict' OR 'innovation climate' OR 'shared perception' OR norms OR values OR 'basic assumptions'
Tension for change	'tension for change' OR 'need for change' OR 'climate for change'
Compatibility	'compatibility of climate' OR 'organizational fit' OR 'organizational fitness' OR 'change required' OR compatibility OR 'change management'
Relative priority	'relative priority' OR 'shared perception*' OR importance
Organizational incentives and rewards	'organizational incentives' OR 'organizational rewards' OR 'extrinsic incentives' OR 'goal-sharing awards' OR 'performance review*' OR promotion OR 'goal sharing'
Goals and feedback	([goals OR 'goals and objectives'] AND feedback) OR 'audit and feedback' OR 'supervisor feedback' OR 'goal* feedback*' OR 'clinical evaluation' OR 'clinical feedback' OR 'performance feedback' OR 'performance evaluation' OR 'economic evaluation'
Learning climate	'learning climate' OR 'learning capability' OR 'peer collaboration' OR 'reflective thinking' OR 'psychological safety' OR 'change management' OR 'learning culture' OR 'professional growth' OR 'professional development'

The most frequently used and highest rated measure of organizational culture in behavioral health (with 46 uses of culture and/or climate scales) was the Organizational Social Context–Culture Scale (Glisson et al., 2008). It received a total score of 11 (maximum possible score = 36) and had evidence of internal consistency ("3—good"), convergent validity ("1—minimal"), predictive validity ("1—minimal"), concurrent validity ("2—adequate"), structural validity ("2—adequate"),

responsiveness ("2—adequate"), and norms ("1—minimal"), along with a "–1—poor" rating for known-groups validity. The next highest scoring measure of organizational culture was the Organizational Description Questionnaire (Parry & Proctor-Thomson, 2001) that was used eight times (total score = 9; maximum possible score = 36), with ratings of "2—adequate" for internal consistency, "3—good" for convergent validity, and "4—excellent" for norms."

Table 3. Inclusion and exclusion criteria.

Domain	From inclusion/exclusion criteria
Intervention	<p>Include:</p> <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, screening <p>Exclude:</p> <ul style="list-style-type: none"> Physical health interventions (e.g., surgery)
Implementation focus	<p>Include:</p> <ul style="list-style-type: none"> Studies demonstrating relevance to implementation, defined as the process of integrating evidence-based practices into a community setting (e.g., a study evaluating organizational capacity for implementing an evidence-based practice) <p>Exclude:</p> <ul style="list-style-type: none"> Studies that do not focus on implementation of an evidence-based practice (e.g., a pure effectiveness trial of an intervention)
Outcomes	<p>Include:</p> <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, social and role functioning <p>Exclude:</p> <ul style="list-style-type: none"> Physical health outcomes (e.g., blood pressure)
Setting	<p>Include:</p> <ul style="list-style-type: none"> Behavioral health settings including but not limited to: mental health treatment centers, medical care facilities in which behavioral health is integrated, criminal justice, education, social service <p>Exclude:</p> <ul style="list-style-type: none"> N/A

Table 4. Definitions of psychometric properties.

Psychometric property	Definition
Internal consistency	Whether several items that purport to measure the same construct actually produce a similar score in the same test (Haynes et al., 1999)
Convergent validity	The degree to which two constructs that are theoretically related are in fact related (Haynes et al., 1999)
Discriminant (or divergent) validity	The degree to which two constructs that are theoretically distinct are in fact distinct (Haynes et al., 1999)
Known-groups validity	The degree to which a measure can distinguish groups with differing characteristics (e.g., those who are clinically depressed from those who are feeling “blue”) (Davidson, 2014)
Structural validity	The degree to which all test items rise or fall together (Ginty, 2013)
Predictive validity	The degree to which a measure can predict or correlate with an outcome of interest measured at some point in the future (Lin & Yao, 2014)
Concurrent validity	The degree to which two measurements taken at the same time correlate, and the measure under consideration is compared to an established measure of the same construct (Haynes et al., 1999)
Responsiveness	The degree to which a measure detects a meaningful change in the construct in measures over time (Haynes et al., 1999)
Norms	Measured by sample size, means, and standard deviations, norms are meant to assess generalizability

Note. See Additional File 2 of Lewis, Mettert, et al. (2018) for the complete rating scale for each psychometric criterion.

Organizational climate

We identified 36 measures of organizational climate, 32 of which are subscales of broader measures (e.g., Survey of Organizational Functioning–Organizational Climate Domain; Broome et al., 2007). Measures were primarily developed in the United States (94%); used more than once (75%); used most frequently in outpatient community mental health (92%) and residential care settings (47%); administered at the provider (78%), director (67%), supervisor (64%), and clinic/site levels (61%);

used within substance use disorder (72%) and general mental health services (50%); and were used most often at the implementation (64%) and exploration phases (53%).

Evidence for internal consistency was available for 31 measures, convergent validity for five measures, discriminant validity for one measure, known-groups validity for 13 measures, predictive validity for 19 measures, concurrent validity for two measures, structural validity for five measures, responsiveness for three measures, and norms for 35 measures.

The median rating for internal consistency was “3—good,” for convergent validity “3—good,” for discriminant validity “1—poor,” for known-groups validity “2—adequate,” for predictive validity “1—minimal,” for concurrent validity “1—minimal,” for structural validity “2—adequate,” for responsiveness “1—minimal,” and for norms “2—adequate.” The median rating of “1—poor” for discriminant validity was based upon one measure: the Organizational Climate Measure (Patterson et al., 2005).

The measure that scored the highest (total score=13; maximum possible score=36) among the organizational climate measures was the Texas Christian University Program Training Needs Survey (Simpson, 2002), which was used five times and showed evidence of internal consistency (“2—adequate”), known-groups validity (3—good”), predictive validity (“2—adequate”), structural validity (“2—adequate”), and norms (“4—excellent”). The Organizational Social Context—Climate (Glisson et al., 2008), the most frequently used measure in behavioral health received a total score of 12 (maximum possible score=36). This included evidence of internal consistency (“3—good”), known-groups validity (“1—minimal”), predictive validity (“2—adequate”), concurrent validity (“1—minimal”), structural validity (“2—adequate”), responsiveness (“1—minimal”), and norms (“2—adequate”). Finally, few measures used in behavioral health focus solely on organizational climate. One exception is the Organizational Climate Measure (Patterson et al., 2005). It had been used twice in behavioral health, had a total score of 9 (maximum possible score=36), and had evidence of internal consistency (“2—adequate”), discriminant validity (“1—poor”), predictive validity (“2—adequate”), concurrent validity (“1—minimal”), structural validity (“1—minimal”), and norms (“4—excellent”).

Implementation climate

For implementation climate, we included measures directly addressing implementation climate or measures of any of the six subconstructs that the CFIR includes as contributing to a positive implementation climate, including tension for change, compatibility, relative priority, organizational incentives and rewards, goals and feedback, and learning climate. We refer readers to Table 5 for descriptive information.

We identified two measures of implementation climate, one of which is a subscale of a broader measure (Readiness for Integrated Care Questionnaire—Implementation Climate Scale; Scott et al., 2017).

Of two measures of implementation climate, evidence of norms was available for both and evidence of internal consistency, convergent validity, discriminant validity, predictive validity, structural validity, and norms was available for one measure. Neither measure had evidence for known-groups validity, concurrent validity, or responsiveness.

The Implementation Climate Scale (Ehrhart, Aarons, et al., 2014) had the highest overall rating (total score = 11;

maximum possible score=36) from five uses in behavioral health, demonstrating evidence of internal consistency (“3—good”), convergent validity (“2—adequate”), discriminant validity (“1—minimal”), predictive validity (“1—minimal”), structural validity (“2—adequate”), and norms (“2—adequate”).

Tension for change. We identified two measures of tension for change, both of which were subscales of broader measures (e.g., Texas Christian University Organizational Readiness for Change—Pressures for Change Scale; Lehman et al., 2002). Evidence of internal consistency and norms was available for both measures, and evidence of convergent validity, known-groups validity, and predictive validity was available for one measure. There was no evidence of discriminant validity, concurrent validity, structural validity, or responsiveness. Both measures were rated the same (total score=3; maximum possible score=36). The Texas Christian University Organizational Readiness for Change—Pressures for Change Subscale (Lehman et al., 2002) demonstrated evidence of internal consistency (“2—adequate”), convergent validity (“1—minimal”), and norms (“2—adequate”); however, both known-groups validity and predictive validity were rated as (“1—poor”) despite being used 37 times in behavioral health. The Survey of Organizational Functioning—Pressures for Change Subscale (Broome et al., 2007) was used 12 times and exhibited evidence of internal consistency (“1—minimal”) and norms (“2—adequate”).

Compatibility. We identified six measures of compatibility, all of which were subscales of broader measures (Perceived Characteristics of Intervention Scale—Compatibility Scale; Cook et al., 2015). Evidence of internal consistency was available for three measures, evidence of predictive validity was available for one measure, and evidence of norms was available for two measures. There was no evidence for convergent validity, discriminant validity, known-groups validity, concurrent validity, structural validity, or responsiveness. The highest rated measure was the Perceived Characteristics of Intervention Scale—Compatibility Subscale (Cook et al., 2015), which had been used twice and received a total score of five (maximum possible score=36) and demonstrated evidence of internal consistency (“3—good”) and norms (“2—adequate”). The next highest rated measure was the Cook Implementation Measure—Compatibility Scale (Cook et al., 2012), which had been used four times and showed evidence of internal consistency (“3—good”) and predictive validity (“1—minimal”).

Relative priority. We identified two measures of relative priority, both of which are subscales of broader measures (e.g., Cook Implementation Measure—Goals and Priorities; Cook et al., 2012). Evidence of internal consistency was available for one measure and evidence of norms was available for one measure. There was no information available

Table 5. Description of measures and subscales.

	Organizational culture (N=21)		Organizational climate (N=36)		Implementation climate (N=2)		Tension for change (N=2)		Compatibility (N=6)		Relative priority (N=2)		Organizational incentives and rewards (N=3)		Goals and feedback (N=3)		Learning climate (N=2)		
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	
One-time use only																			
Yes	5	24	9	25	1	50	0	0	4	67	1	50	1	33	1	33	1	50	
No	16	76	27	75	1	50	2	100	2	33	1	50	2	67	2	67	1	50	
Country																			
United States	20	95	34	94	2	100	2	100	5	83	2	100	3	100	3	100	2	100	
Other	1	5	2	6	0	0	0	0	1	17	0	0	0	0	0	0	0	0	
Setting ^a																			
State mental health	1	5	0	0	1	50	0	0	0	0	0	0	1	33	0	0	0	0	
Inpatient psychiatry	6	29	8	22	1	50	1	50	0	0	0	0	1	33	0	0	1	50	
Outpatient community	18	86	33	92	2	100	2	100	2	33	1	50	2	67	1	33	2	100	
School mental health	1	5	6	17	0	0	1	50	0	0	0	0	2	67	0	0	0	0	
Residential care	15	71	17	47	0	0	1	50	2	33	1	50	2	67	1	33	2	100	
Other	14	67	31	86	2	100	2	100	3	50	1	50	2	67	2	67	1	50	
Level ^a																			
Consumer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Organization	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Clinic/site	8	38	22	61	1	50	2	100	0	0	0	0	1	33	1	33	1	50	
Provider	21	100	28	78	2	100	2	100	5	83	2	100	3	100	3	100	2	100	
System	1	5	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Team	2	10	0	0	1	50	0	0	0	0	0	0	1	33	1	33	0	0	
Director	8	38	24	67	1	50	2	100	2	33	2	100	1	33	2	67	1	50	
Supervisor	11	52	23	64	2	100	2	100	1	17	1	50	1	33	1	33	1	50	
Other	7	33	1	3	0	0	0	0	1	17	0	0	1	33	0	0	0	0	
Population ^a																			
General mental health	18	86	18	50	2	100	1	50	2	33	1	50	2	67	1	33	1	50	
Anxiety	0	0	6	17	1	50	1	50	0	0	0	0	1	33	0	0	0	0	
Depression	3	14	14	39	0	0	1	50	0	0	0	0	0	0	1	33	0	0	
Suicidal ideation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Alcohol use disorder	12	57	9	25	0	0	1	50	1	17	0	0	0	0	0	0	1	50	
Substance use disorder	14	67	26	72	0	0	1	50	1	17	0	0	0	0	0	0	2	100	
Behavioral disorder	0	0	7	19	0	0	1	50	0	0	0	0	0	0	0	0	0	0	
Mania	7	33	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Eating disorder	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grief	0	0	0	0	0	0	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	0	0	0	0	
Trauma	0	0	1	3	0	0	0	0	2	33	1	50	1	33	1	33	0	0	
Other	0	0	6	17	1	50	1	50	1	17	0	0	1	33	0	0	0	0	
Phase of implementation ^a																			
Exploration	15	71	19	53	1	50	2	100	2	33	2	100	1	33	1	33	1	50	
Preparation	3	14	7	19	0	0	1	50	0	0	0	0	0	0	0	0	0	0	
Implementation	9	43	23	64	1	50	2	100	3	50	1	50	3	100	2	67	1	50	
Sustainment	1	5	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Not specified	1	5	9	25	0	0	0	0	2	33	0	0	0	0	1	33	1	50	

^aSome measures were used in multiple settings, levels, and phases of implementation.

Table 6. Psychometric information availability.

	Organizational culture (N=21)		Organizational climate (N=36)		Implementation climate (N=2)		Tension for change (N=2)		Compatibility (N=6)		Relative priority (N=2)		Organizational incentives and rewards (N=3)		Goals and feedback (N=3)		Learning climate (N=2)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Internal consistency	18	86	31	86	1	50	2	100	3	50	1	50	3	100	2	67	1	50
Convergent validity	2	10	5	14	1	50	1	50	0	0	0	0	0	0	1	33	1	50
Discriminant validity	0	0	1	3	1	50	0	0	0	0	0	0	0	0	0	0	0	0
Known-groups validity	2	10	13	36	0	0	1	50	0	0	0	0	0	0	0	0	0	0
Predictive validity	12	57	19	53	1	50	1	50	1	17	0	0	1	33	1	33	1	50
Concurrent validity	2	10	2	6	0	0	0	0	0	0	0	0	0	0	0	0	1	50
Structural validity	1	5	5	14	1	50	0	0	0	0	0	0	0	0	0	0	0	0
Responsiveness	1	5	3	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Norms	20	95	35	97	2	100	2	100	2	33	1	50	1	33	1	33	2	100

Table 7. Summary statistics for instrument ratings.

	Organizational culture (N=21)			Organizational climate (N=36)			Implementation climate (N=2)			Tension for change (N=2)			Compatibility (N=6)			Relative priority (N=2)			Organizational incentives and rewards (N=3)			Goals and feedback (N=3)			Learning climate (N=2)		
	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range	Median	Range	Range
Internal consistency	2	-1,4	3	-1,4	3	-	1	1,2	3	-	3	-	2	-	3	-	2	-	3	-	4	-	-	-	-	-	-
Convergent validity	2	1,3	3	2,3	1	-	1	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-
Discriminant validity	-	-	-1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Known-groups validity	-1	-	2	-1,4	-	-	-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Predictive validity	1	-1,1	1	-1,3	1	-	-1	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Concurrent validity	-1	-1,2	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Structural validity	2	-	2	-1,3	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Responsiveness	-	-	1	-1,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Norms	2	-1,4	2	-1,4	-1	-1,2	2	-	-1	-1,2	-1	-	2	-	2	-	2	-	1	-	2	-	-	-	-	-	-

Note. Scores for individual measures ranged from -1 ("poor") to 4 ("excellent"). Median, excluding zeros where psychometric information was not available. When the median of two scores would equal "0" (e.g., a score of -1 and 1), the lower score was taken.

Table 8. Psychometric ratings for each measure by focal construct.

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Organizational culture (N=21)										
Group Innovation Inventory (Caldwell & O'Reilly, 2003)	3	0	0	0	0	-1	0	0	3	5
Organizational Culture Profile (Dark et al., 2015)	0	0	0	0	0	0	0	0	1	1
Organizational Description Questionnaire (Parry & Proctor-Thomson, 2001)	2	3	0	0	0	0	0	0	4	9
Organizational Readiness for Change Assessment-Staff Culture Subscale (Helfrich et al., 2009)	4	0	0	-1	0	0	0	0	-1	2
Organizational Social Context Survey-Culture Scales (Glisson et al., 2008)	3	1	0	-1	1	2	2	2	1	11
Readiness for Integrated Care Questionnaire-Culture Subscale (Scott et al., 2017)	0	0	0	0	0	0	0	0	-1	-1
Rubenstein Implementation Capability Composites-Depression Culture & Attitudes (Rubenstein et al., 2014)	2	0	0	0	-1	0	0	0	1	2
Rubenstein Implementation Capability Composites-Quality Improvement Culture & Attitudes (Rubenstein et al., 2014)	2	0	0	0	-1	0	0	0	1	2
Systems of Care Implementation Survey-Values & Principles Subscale (Boothroyd et al., 2011)	1	0	0	0	0	0	0	0	0	1
The National Criminal Justice Treatment Practices Survey-Cohesive Culture Subscale (Taxman et al., 2007)	0	0	0	0	1	0	0	0	2	3
The National Criminal Justice Treatment Practices Survey-Hierarchical Consistency Subscale (Taxman et al., 2007)	2	0	0	0	-1	0	0	0	1	2
The National Criminal Justice Treatment Practices Survey-Organizational Culture Domain (Taxman et al., 2007)	2	0	0	0	-1	0	0	0	2	3
The National Criminal Justice Treatment Practices Survey-Organizational Culture Domain-Staff Influence on Treatment Improvement Subscale (Taxman et al., 2007)	4	0	0	0	0	0	0	0	2	6
The National Criminal Justice Treatment Practices Survey-Organizational Culture Domain-Importance of Other Services Relative to Drug Abuse Treatment Subscale (Taxman et al., 2007)	4	0	0	0	0	0	0	0	2	6
The National Criminal Justice Treatment Practices Survey-Organizational Culture Domain-Correctional Staff Respect for Treatment (Taxman et al., 2007)	3	0	0	0	-1	0	0	0	2	4
Work Environment Scale-Autonomous Subscale (Moos et al., 1982)	1	0	0	0	1	0	0	0	2	4
Work Environment Scale-Clarity Subscale (Moos et al., 1982)	1	0	0	0	1	0	0	0	2	4
Work Environment Scale-Control Subscale (Moos et al., 1982)	-1	0	0	0	0	0	0	0	1	0
Work Environment Scale-Peer Cohesion Subscale (Moos et al., 1982)	1	0	0	0	1	0	0	0	2	4
Work Environment Scale-Task Orientation Subscale (Moos et al., 1982)	2	0	0	0	1	0	0	0	2	5
Work Environment Scale-Work Pressure Subscale (Moos et al., 1982)	-1	0	0	0	1	0	0	0	2	2

(Continued)

Table 8. (Continued)

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Organizational climate (N=36)										
Clinical Practice Organizational Survey (CPOS)-2007 Survey-Competing Demands and Stress Subscale (Chang et al., 2013)	3	0	0	-1	-1	0	0	0	2	3
Clinical Practice Organizational Survey (CPOS)-2007 Survey-External Authority Over Relationship with Subspecialists Subscale (Chang et al., 2013)	3	0	0	-1	-1	0	0	0	2	3
Clinical Practice Organizational Survey (CPOS)-2007 Survey-Internal Authority Over Primary Care Clinic Subscale (Chang et al., 2013)	4	0	0	-1	-1	0	0	0	2	4
Clinical Practice Organizational Survey (CPOS)-2007 Survey-Organizational Climate Domain (Chang et al., 2013)	3	0	0	-1	-1	0	0	0	2	3
Clinical Practice Organizational Survey (CPOS)-2007 Survey-Oriented Toward Quality Improvement Subscale (Chang et al., 2013)	3	0	0	2	1	0	0	0	2	8
Clinical Practice Organizational Survey (CPOS)-2007 Survey-Resistance Subscale (Chang et al., 2013)	4	0	0	-1	-1	0	0	0	2	4
Cook Implementation Measure-Risk Taking Climate (Cook et al., 2012)	3	0	0	0	0	0	0	0	0	3
Organizational Readiness for Change Scale-Director-Climate Domain (Lehman et al., 2002)	0	0	0	0	0	0	0	0	2	2
Organizational Social Context Survey-Climate Scales (Glisson et al., 2008)	3	0	0	1	2	1	2	1	2	12
Readiness for Integrated Care Questionnaire-Climate Subscale (Scott et al., 2017)	0	0	0	0	0	0	0	0	-1	-1
Survey of Organizational Functioning-Organizational Climate Domain (Broome et al., 2007)	0	0	0	0	1	0	-1	0	3	3
Survey of Organizational Functioning-Autonomy Subscale (Broome et al., 2007)	1	0	0	0	0	0	0	0	2	3
Survey of Organizational Functioning-Cohesion Subscale (Broome et al., 2007)	3	0	0	0	0	0	0	0	2	5
Survey of Organizational Functioning-Collective Responsibility Subscale (Broome et al., 2007)	3	0	0	0	0	0	0	0	2	5
Survey of Organizational Functioning-Mission Subscale (Broome et al., 2007)	2	0	0	0	0	0	0	0	1	3
Survey of Organizational Functioning-Openness to Change Subscale (Broome et al., 2007)	2	0	0	0	0	0	0	0	2	4
Survey of Organizational Functioning-Stress Subscale (Broome et al., 2007)	3	0	0	0	3	0	0	0	2	8
TCU Program Training Needs Survey (Simpson, 2002)	2	0	0	3	2	0	2	0	4	13
Team Assessment Questionnaire-Team Climate and Atmosphere Subscale (Mahoney et al., 2012)	0	0	0	0	0	0	0	0	2	2
Team Climate Inventory (Anderson & West, 1998)	4	0	0	0	0	0	3	0	2	9
Team Fitness Test (Chodosh et al., 2015)	0	0	0	0	0	0	0	3	-1	2

(Continued)

Table 8. (Continued)

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Texas Christian University Organizational Readiness for Change—Organizational Climate Domain (Lehman et al., 2002)	1	0	0	0	1	0	0	-1	2	3
Texas Christian University Organizational Readiness for Change—Autonomy Subscale (Lehman et al., 2002)	-1	2	0	2	1	0	0	0	2	6
Texas Christian University Organizational Readiness for Change—Change Subscale (Lehman et al., 2002)	2	2	0	2	1	0	0	0	2	9
Texas Christian University Organizational Readiness for Change—Cohesion Subscale (Lehman et al., 2002)	3	3	0	2	1	0	0	0	2	11
Texas Christian University Organizational Readiness for Change—Mission Subscale (Lehman et al., 2002)	2	3	0	4	1	0	0	0	2	12
Texas Christian University Organizational Readiness for Change—Stress Subscale (Lehman et al., 2002)	2	3	0	2	1	0	0	0	3	11
The Children's Services Survey—Depersonalization Subscale (Glisson & Hemmelgarn, 1998)	3	0	0	0	0	0	0	0	2	5
The Children's Services Survey—Emotional Exhaustion Subscale (Glisson & Hemmelgarn, 1998)	4	0	0	0	0	0	0	0	2	6
The Children's Services Survey—Fairness Subscale (Glisson & Hemmelgarn, 1998)	1	0	0	0	0	0	0	0	2	3
The Children's Services Survey—Growth & Advancement Subscale (Glisson & Hemmelgarn, 1998)	3	0	0	0	0	0	0	0	2	5
The Children's Services Survey—Role Clarity Subscale (Glisson & Hemmelgarn, 1998)	3	0	0	0	0	0	0	0	2	5
The Children's Services Survey—Role Conflict Subscale (Glisson & Hemmelgarn, 1998)	3	0	0	0	0	0	0	0	2	5
The National Criminal Justice Treatment Practices Survey—Organizational Climate Subscale (Taxman et al., 2007)	4	0	0	0	1	0	0	0	3	8
The Organizational Climate Measure (Patterson et al., 2005)	2	0	-1	0	2	1	1	0	4	9
Work Environment Scale—Innovation Subscale (Moos et al., 1982)	3	0	0	0	1	0	0	0	2	6
Implementation climate (N=2)										
Implementation Climate Scale (Ehrhart, Aarons, et al., 2014)	3	2	1	0	1	0	2	0	2	11
Readiness for Integrated Care Questionnaire—Implementation Climate Support Subscale (Scott et al., 2017)	0	0	0	0	0	0	0	0	-1	-1
Tension for change (N=2)										
Survey of Organizational Functioning—Pressures for Change Subscale (Broome et al., 2007)	1	0	0	0	0	0	0	0	2	3
Texas Christian University Organizational Readiness for Change—Pressures for Change Subscale (Lehman et al., 2002)	2	1	0	-1	-1	0	0	0	2	3
Compatibility (N=6)										
Cook Implementation Measure—Compatibility (Cook et al., 2012)	3	0	0	0	1	0	0	0	0	4

(Continued)

Table 8. (Continued)

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Malte Post-Treatment Smoking Cessation Beliefs Measure–Compatibility (Malte et al., 2013)	0	0	0	0	0	0	0	0	0	0
Moise Attributes of Innovation Adoption–Compatibility Subscale (Moise et al., 2014)	0	0	0	0	0	0	0	0	0	0
Moore & Benbasat Adoption of Information Technology Innovation Measure–Compatibility (Moore & Benbasat, 1991)	3	0	0	0	0	0	0	0	0	3
Perceived Characteristics of Intervention Scale–Compatibility Subscale (Cook et al., 2015)	3	0	0	0	0	0	0	0	2	5
Readiness for Integrated Care Questionnaire–Compatibility/alignment Subscale (Scott et al., 2017)	0	0	0	0	0	0	0	0	–1	–1
Relative priority (N=2)										
Cook Implementation Measure–Goals and Priorities (Cook et al., 2012)	3	0	0	0	0	0	0	0	0	3
Readiness for Integrated Care Questionnaire–Priority Subscale (Scott et al., 2017)	0	0	0	0	0	0	0	0	–1	–1
Organizational incentives and rewards (N=3)										
Cook Implementation Measure–Incentives and Mandates (Cook et al., 2012)	2	0	0	0	0	0	0	0	0	2
Implementation Climate Scale–Rewards Subscale (Ehrhart, Aarons, et al., 2014)	2	0	0	0	1	0	0	0	2	5
Systems of Care Implementation Survey–Provider Accountability Subscale (Boothroyd et al., 2011)	2	0	0	0	0	0	0	0	0	2
Goals and feedback (N=3)										
Chou Measure of Guidelines Information–Feedback Subscale (Chou et al., 2011)	0	2	0	0	–1	0	0	0	0	1
Cook Implementation Measure–Goals and Priorities (Cook et al., 2012)	3	0	0	0	0	0	0	0	0	3
Organizational readiness for Change Assessment–Project Progress Tracking Subscale (Helfrich et al., 2009)	3	0	0	0	0	0	0	0	1	4
Learning climate (N=2)										
Ramsey Learning Climate Measure (Ramsey et al., 2015)	4	–1	0	0	0	1	0	0	2	6
The National Criminal Justice Treatment Practices Survey–Climate for Learning Subscale (Taxman et al., 2007)	0	0	0	0	–1	0	0	0	2	1

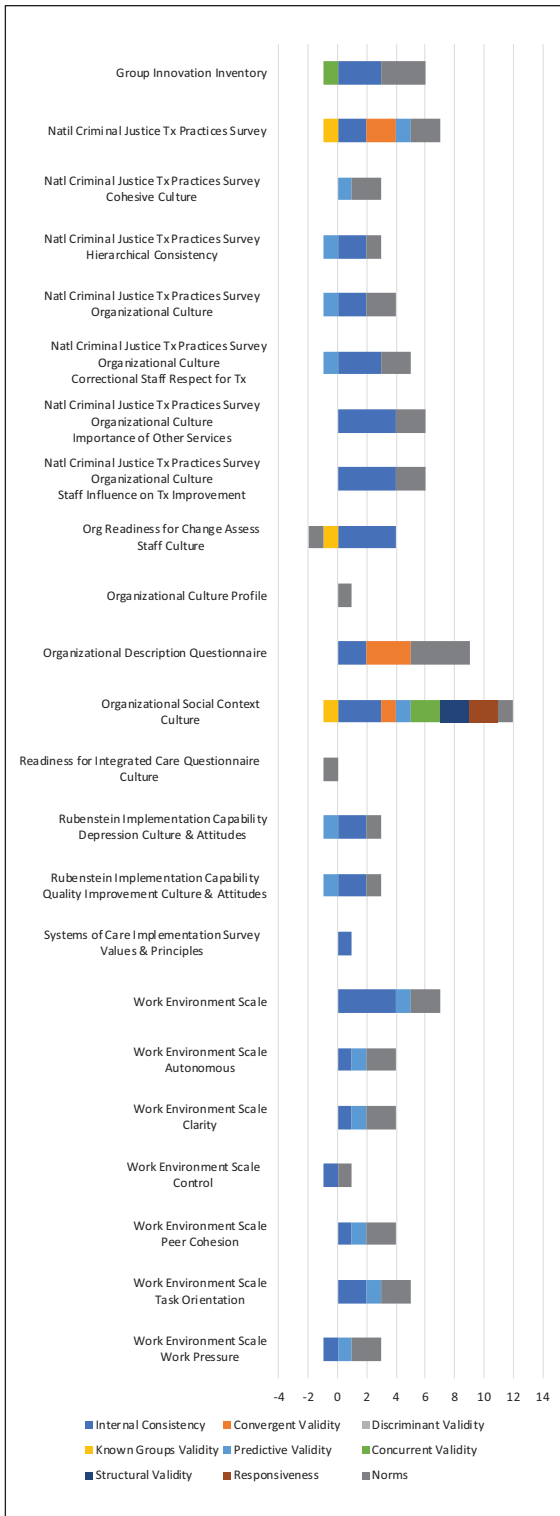


Figure 1. Head-to-head comparison of measures of organizational culture.

on any of the remaining psychometric criteria. The highest rated measure was the Cook Implementation Measure–Goals and Priorities (Cook et al., 2012), which had been used four times and a total score of three (maximum

possible score=36) based upon evidence of internal consistency (“3—good”).

Organizational incentives and rewards. We identified three measures of organizational incentives and rewards, all of which were subscales of broader measures (e.g., Implementation Climate Scale–Rewards Scale; Ehrhart, Aarons, et al., 2014). Evidence of internal consistency was available for all three measures, and evidence of predictive validity and norms was available for one measure. No further information about psychometric properties was available. The Implementation Climate Scale–Rewards Subscale (Ehrhart, Aarons, et al., 2014) was used five times and received the highest overall rating (total score=5; maximum possible score=36), demonstrating evidence of internal consistency (“2—adequate”), predictive validity (“1—minimal”), and norms (“2—adequate”).

Goals and feedback. We identified three measures of goals and feedback, all of which were subsets of broader measures (e.g., Chou Measure of Guideline Information–Feedback Scale; Chou et al., 2011). Evidence for internal consistency was available for two measures, and evidence of convergent validity, predictive validity, and norms was available for one measure. No other information on psychometric properties was available. The Organizational Readiness for Change Assessment–Project Progress Tracking Subscale (Helfrich et al., 2009; four uses in behavioral health) was rated the highest (total score=4; maximum possible score=36), with evidence of internal consistency (“3—good”) and norms (“1—minimal”). The Cook Implementation Measure–Goals and Priorities Subscale (Cook et al., 2012) received a total score of three, with evidence of internal consistency (“3—good”).

Learning climate. We identified two measures of learning climate, one of which was a subscale from a broader measure (The National Criminal Justice Treatment Practices Survey–Climate for Learning Scale; Taxman et al., 2007). Evidence of norms was available for two measures and evidence for internal consistency, convergent validity, predictive validity, and concurrent validity were available for one measure. There was no evidence of discriminant validity, known-groups validity, structural validity, or responsiveness. The Ramsey Learning Climate Measure (Ramsey et al., 2015) was rated the highest (total score=6; maximum possible score=36), with evidence of internal consistency (“4—excellent”), convergent validity (“–1—poor”), concurrent validity (“1—minimal”), and norms (“2—adequate”).

Discussion

Summary of findings

This systematic review of measures of organizational culture, organizational climate, implementation climate, and

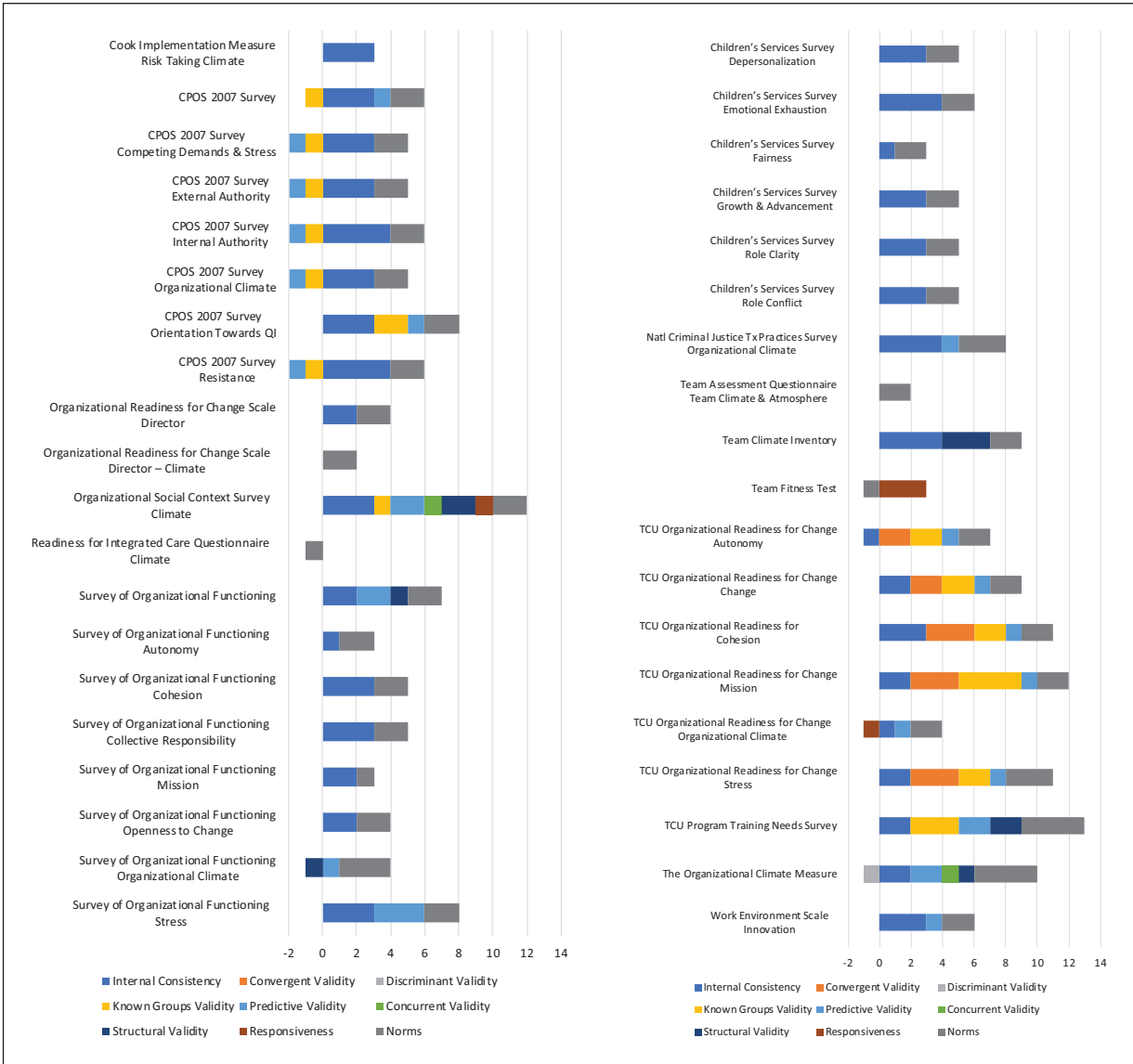


Figure 2. Head-to-head comparison of measures of organizational climate.

related constructs in behavioral health identified some promising measures; however, consistent with other reviews of organizational constructs (Allen et al., 2017; Clinton-McHarg et al., 2016; Weiner et al., 2020), the overall state of measurement across these constructs is poor. While 21 measures of organizational culture and 36 measures of organizational climate were identified, the vast majority were subscales within broader measures. Far fewer measures of implementation climate and related constructs were identified. Previous work has documented the problem of “home-grown” measures that are used only once (Lewis et al., 2015; Martinez et al., 2014). Encouragingly, more than 75% of measures of organizational culture and organizational climate identified in this review were used more than once, which may reflect the long tradition of these constructs in the broader literature

(Ehrhart, Schneider, et al., 2014). In contrast, nearly half of the measures of implementation climate and related sub-constructs were used only once, perhaps reflecting its more recent emergence in the field (Klein & Sorra, 1996; Weiner et al., 2011).

Limited psychometric evidence was available for the identified measures of organizational culture, organizational climate, implementation climate, and its sub-constructs. This is consistent with findings from previous reviews of a broader set of implementation constructs (Chaudoir et al., 2013; Clinton-McHarg et al., 2016), as well as findings from a recent review of organizational readiness for change (Weiner et al., 2020). For organizational culture and organizational climate, evidence of internal consistency and norms was available for most measures. Evidence of predictive validity was available

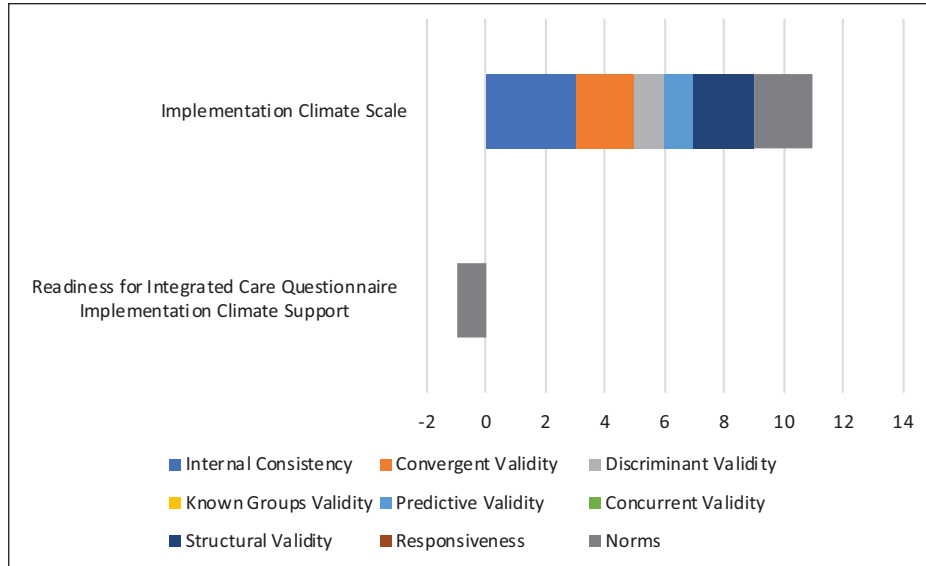


Figure 3. Head-to-head comparison of measures of implementation climate.

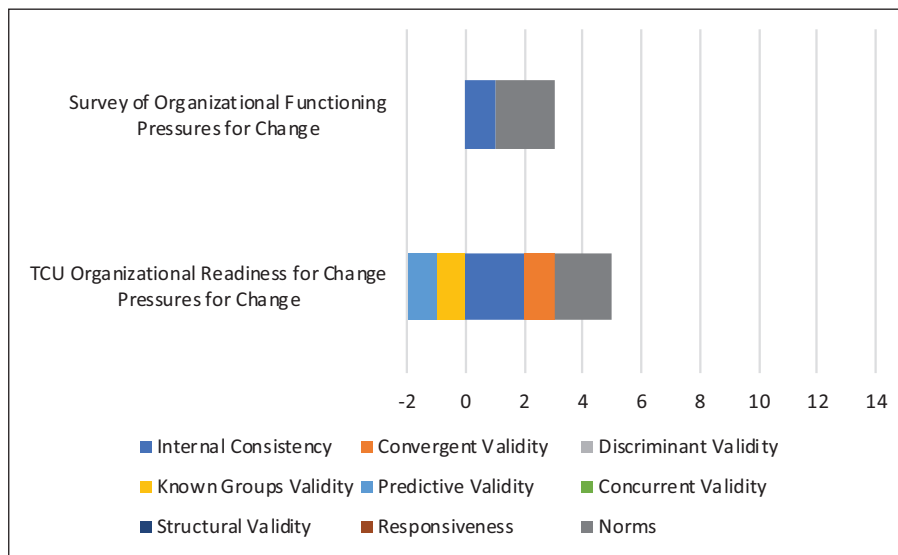


Figure 4. Head-to-head comparison of measures of tension for change.

for over half of identified measures, though nine of them received a rating of “poor” suggesting that evidence did not support study hypotheses. Evidence for other psychometric properties like known-groups validity, concurrent validity, convergent validity, discriminant validity, structural validity, and responsiveness was sparse. Generally, psychometric evidence for implementation climate and its related subconstructs was less readily available. Only one measure of organizational culture (Glisson et al., 2008), five measures of organizational climate (Anderson & West, 1998; Broome et al., 2007; Glisson et al., 2008; Patterson et al., 2005; Simpson, 2002), and one measure of implementation climate (Ehrhart, Aarons, et al., 2014)

were assessed for structural validity, which is concerning given that a measure’s dimensionality should be checked prior to checking its internal consistency (DeVellis, 2012). Also concerning is a striking lack of evidence for measure responsiveness (i.e., sensitivity to change), as only four measures among all focal constructs possessed evidence of responsiveness (Chodosh et al., 2015; Glisson et al., 2008; Lehman et al., 2002). This weakness will stymie efforts to identify organizational-level mechanisms that explain how and why implementation strategies can improve implementation and clinical outcomes (Lewis et al., 2020; Lewis, Klasnja, et al., 2018; Williams, 2016; Williams et al., 2017).

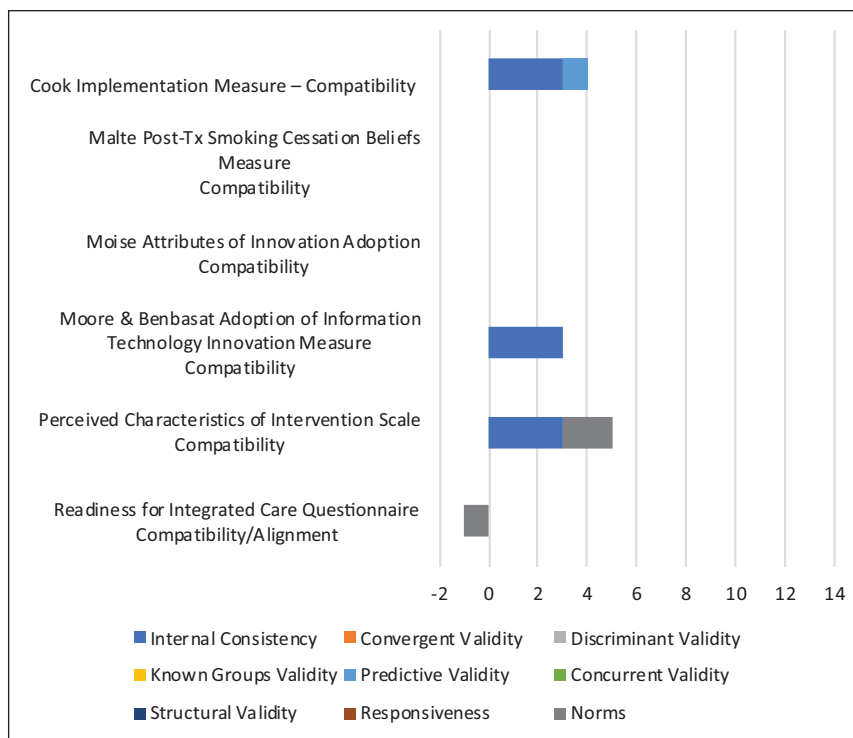


Figure 5. Head-to-head comparison of measures of compatibility.

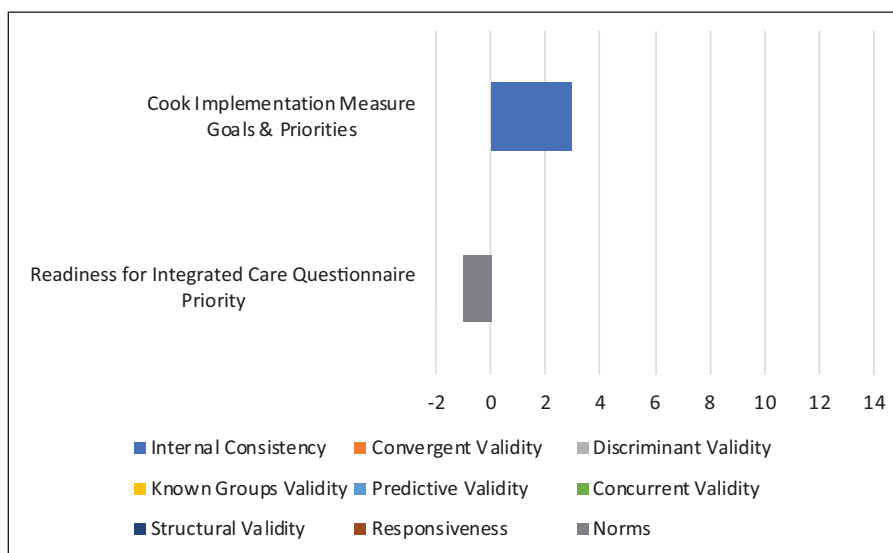


Figure 6. Head-to-head comparison of measures of relative priority.

Overall measurement quality was found to be poor. With the exception of internal consistency, most median ratings ranged from “–1—poor” to “2—adequate.” Only seven measures received an overall score of 10 or higher (out of a possible score of 36) on the PAPERS psychometric rating criteria (Lewis, Mettert, et al., 2018; Stanick et al., 2021). The Organizational Social Context measures of culture and climate received scores of 11 and 12, respectively, and represent

the most frequently studied measure in behavioral health-focused implementation research with national norms established in mental health (Glisson et al., 2008) and child welfare (Glisson et al., 2012). An additional four measures were in the organizational climate domain, including the Texas Christian University Program Training Needs Survey (total score=13; Simpson, 2002) and three subscales from the Texas Christian University Organizational Readiness for Change measure

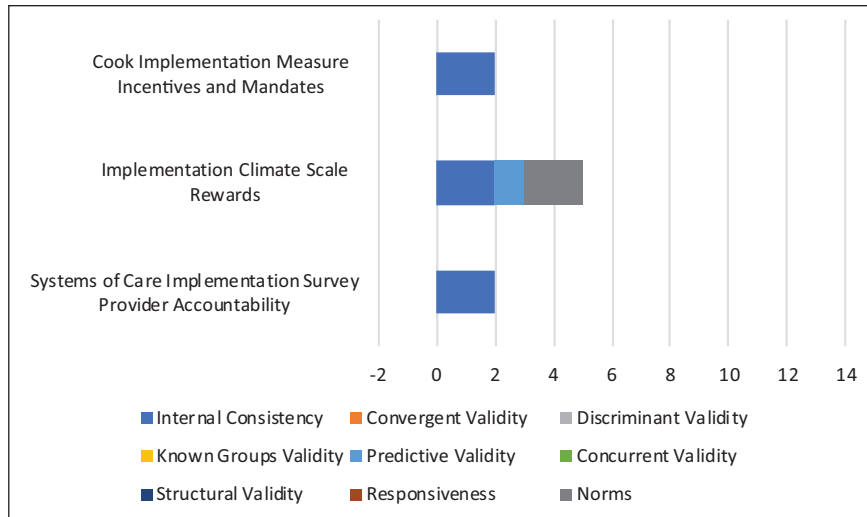


Figure 7. Head-to-head comparison of measures of organizational incentives and rewards.

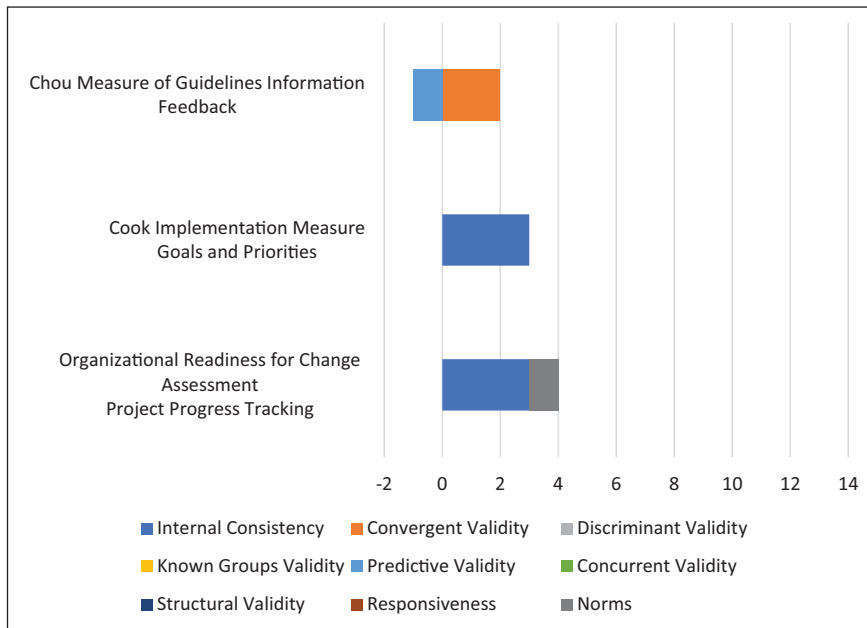


Figure 8. Head-to-head comparison of measures of goals and feedback.

(“Mission,” “Cohesion,” and “Stress”; Lehman et al., 2002) that total scores of 12, 11, and 11. The Texas Christian University Program Training Needs Survey (Simpson, 2002) has only been used five times in behavioral health, but with ratings of “2—adequate” to “4—excellent” on five different psychometric criteria, it may have promise for further use and evaluation. While the Texas Christian University Organizational Readiness for Change measure (Lehman et al., 2002) scored relatively high in comparison to other measures included in this review, there was no evidence of structural validity or responsiveness, and only “minimal” evidence of predictive validity despite 37 uses in behavioral

health, suggesting that more uses may not offer more positive psychometric evidence. The last measure to receive a score of 10 or higher was the Implementation Climate Scale (total score=11; Ehrhart, Aarons, et al., 2014), which has been used five times in behavioral health. Given its promising psychometric properties and desirable pragmatic properties (free, only 18 items), this scale demonstrates promise.

Future directions

There is a need to prioritize further psychometric evaluation of promising measures that have yet been used frequently in

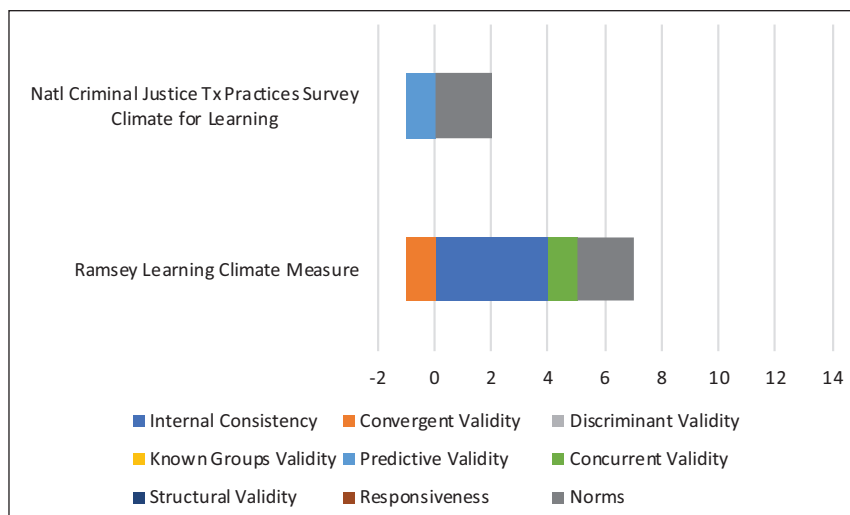


Figure 9. Head-to-head comparison of measures of learning climate.

behavioral health. There are also opportunities to rigorously develop new measures of sparsely populated constructs, particularly for the subconstructs of measures of implementation climate.

Though we did not explicitly consider the extent to which identified measures are pragmatic (Powell et al., 2017; Stanick et al., 2018, 2021), it will be critical to do so moving forward. Some measures identified in this review are brief and freely available, while others are quite long and proprietary. Measures' pragmatic properties are likely to influence their use in both research and applied implementation efforts.

Organizational culture and implementation climate are broad constructs that have been conceptualized and measured in a wide range of ways (Aarons et al., 2018; Ehrhart, Schneider, et al., 2014; Kimberly & Cook, 2008; Schneider et al., 2013; Verbeke et al., 1998). It would be useful to pursue conceptual and measurement work to delineate ways in which organizational culture and organizational climate have been measured. This work could guide stakeholders wanting to measure specific aspects of organizational culture and organizational climate and illustrate the trade-offs in prioritizing one conceptualization versus another. An additional opportunity may be to develop more holistic profiles of organizational culture and climate using latent profile analysis (Glisson et al., 2014; Williams et al., 2019). For example, Williams et al. (2019) demonstrated that when individual dimensions of culture and climate or the linear combination of all six dimensions were not predictive of fidelity to an EBP, a "comprehensive" profile (high proficiency culture, positive climate) was predictive of fidelity for two of three EBPs. This demonstrates that culture and climate may interact in complex ways, and that "the overall gestalt of the social context may be more important than the level of a single dimension" (Williams et al., 2019, p. 10).

Given calls for improved reporting in implementation research (Wilson et al., 2017), it may be useful to develop reporting guidelines for measurement in implementation studies. These may differ depending upon the type of study. For example, a measure development study may require different minimum criteria as compared to the use of a measure within a broader implementation study.

Limitations

This study has several limitations. First, as with all systematic reviews, it is possible that we failed to identify articles that could have detailed measures of the focal constructs or provided further data on their psychometric evidence. There are at least four potential reasons for this: (1) we did not search explicitly for molar organizational climate since that construct is not included in the CFIR, which was used to generate our search strategy for organizational culture, implementation climate, and related constructs; (2) we did not search the gray literature; (3) the original literature searches for this study were completed in 2017; and (4) we did not search all potentially relevant databases (e.g., PsycINFO, Google Scholar; Bramer et al., 2017). Additional measures of the focal constructs may have been published since the original search date; however, we captured more recent uses of the measures we identified in 2017 by conducting measure-forward "cited-by" searches in May of 2019. Nevertheless, there are also studies that provide additional evidence for included measures that have been published since our measure-forward search (e.g., Beidas et al., 2019; Williams et al., 2020). One measure of implementation climate developed by Jacobs et al. (2014) was not identified in this review (likely because initial development in testing was in both non-behavioral health and behavioral health settings), but appears to

have promising psychometric and pragmatic properties. Second, there are inevitable measures of the focal constructs developed outside of behavioral health, and some of the measures identified in this review may have evidence of further use outside of implementation efforts in behavioral health service settings. Thus, it is important that readers interpret these ratings within this context rather than as an indicator of the measures' overall quality or psychometric strength. Third, it is possible that our assignment of measures and/or subscales to the nine focal constructs was imperfect, particularly given the substantial overlap between the conceptualization and measurement of organizational culture, organizational climate, and related constructs (Kimberly & Cook, 2008). Finally, it is possible that poor reporting practices limit the extent to which evidence was available for identified measures (i.e., it is possible that more thorough evaluations of psychometric properties were conducted but not reported).

Conclusion

This systematic review identifies measures of organizational culture, organizational climate, and implementation climate used in behavioral health-focused implementation studies. Several promising measures were identified, and can inform researchers, EBP purveyors, implementation support practitioners, and others who wish to measure these constructs. However, to enhance understanding of how these constructs influence EBP implementation, there is a need for further testing of the most promising approaches, development of additional psychometrically and pragmatically strong measures, and approaches that elucidate the ways in which “dimensions of organizational culture and climate interact with, reinforce, and counteract one another in complex, non-linear ways as they relate to EBP implementation . . .” (Williams et al., 2019, p. 10).


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 C. Lewis as principal investigator. Dr Lewis is an author of this article and 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) through R01MH106510 (Lewis, PI). Byron Powell was also supported by the NIMH through K01MH113806 (Powell, PI).

ORCID iDs

Byron J Powell  <https://orcid.org/0000-0001-5245-1186>
 Kayne D Mettert  <https://orcid.org/0000-0003-1750-7863>
 Cameo F Stanick  <https://orcid.org/0000-0002-6076-3726>
 Gregory A Aarons  <https://orcid.org/0000-0001-8969-5002>
 Cara C Lewis  <https://orcid.org/0000-0001-8920-8075>

References

- Aarons, G. A., Ehrhart, M. G., & Farahnak, L. R. (2014). Aligning leadership across systems and organizations to develop a strategic climate for evidence-based practice implementation. *Annual Review of Public Health, 35*, 255–274. <https://doi.org/10.1146/annurev-publhealth-032013-182447>
- 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*, 4–23. <https://doi.org/10.1007/s10488-010-0327-7>
- Aarons, G. A., Moullin, J. C., & Ehrhart, M. G. (2018). The role of organizational processes in dissemination and implementation research. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (2nd ed., pp. 121–142). Oxford University Press.
- Albers, B., Metz, A., & Burke, K. (2020). Implementation support practitioners—A proposal for consolidating a diverse evidence base. *BMC Health Services Research, 20*, Article 368. <https://doi.org/10.1186/s12913-020-05145-1>
- Allen, J. D., Towne, S. D., Maxwell, A. E., DiMartino, L., Leyva, B., Bowen, D. J., Linnan, L., & Weiner, B. J. (2017). Measures of organizational characteristics associated with adoption and/or implementation of innovations: A systematic review. *BMC Health Services Research, 17*, Article 591. <https://doi.org/10.1186/s12913-017-2459-x>
- Anderson, N. R., & West, M. A. (1998). Measuring climate for work group innovation: Development and validation of the team climate inventory. *Journal of Organizational Behavior, 19*(3), 235–258.
- Beidas, R. S., Williams, N. J., Becker-Haimes, E. M., Aarons, G. A., Barg, F. K., Evans, A. C., Jackson, K., Jones, D., Hadley, T., Hoagwood, K., Marcus, S. C., Neimark, G., Rubin, R. M., Schoenwald, S. K., Adams, D. R., Walsch, L. M., Zentgraf, K., & Mandell, D. S. (2019). A repeated cross-sectional study of clinicians' use of psychotherapy techniques during 5 years of a system-wide effort to implement evidence-based practices in Philadelphia. *Implementation Science, 14*, Article 67. <https://doi.org/10.1186/s13012-019-0912-4>
- 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). *Journal of Behavioral Health Services & Research, 38*(3), 288–302. <https://doi.org/10.1007/s11414-011-9239-x>
- Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews, 6*, Article 245, 3481–3482. <https://doi.org/10.1186/s13643-017-0644-y>

- Brennan, S. E., Bosch, M., Buchan, H., & Green, S. E. (2012). Measuring organizational and individual factors thought to influence the success of quality improvement in primary care: A systematic review of instruments. *Implementation Science*, 7(121), 1–19. <https://doi.org/10.1186/1748-5908-7-121>
- Broome, K. M., Flynn, P. M., Knight, D. K., & Simpson, D. D. (2007). Program structure, staff perceptions, and client engagement in treatment. *Journal of Substance Abuse Treatment*, 33(2), 149–158. <https://doi.org/10.1016/j.josat.2006.12.030>
- Caldwell, D. F., & O'Reilly, C. A. (2003). The determinants of team-based innovation in organizations: The role of social influence. *Small Group Research*, 34(4), 497–517.
- Chang, E. T., Rose, D. E., Yano, E. M., Wells, K. B., Metzger, M. E., Post, E. P., Lee, M. L., & Rubenstein, L. V. (2013). Determinants of readiness for primary care-mental health integration (PC-MHI) in the VA Health Care System. *Journal of General Internal Medicine*, 28(3), 353–362. <https://doi.org/10.1007/s11606-012-2217-z>
- 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, Article 22. <https://doi.org/10.1186/1748-5908-8-22>
- Chodosh, J., Price, R. M., Cadogan, M. P., Damron-Rodriguez, J., Osterweil, D., Czerwinski, A., Tan, Z. S., Merkin, S. S., Gans, D., & Frank, J. C. (2015). A practice improvement education program using a mentored approach to improving 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>
- 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>
- Cook, J. M., Thompson, R., & Schnurr, P. P. (2015). Perceived characteristics of intervention scale: Development and psychometric properties. *Assessment*, 22(6), 704–714. <https://doi.org/10.1177/1073191114561254>
- 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, Article 50.
- Dark, F., Whiteford, H., Ashkanasy, N. M., Harvey, C., Crompton, D., & Newman, E. (2015). Implementing cognitive therapies into routine psychosis care: Organisational foundations. *BMC Health Services Research*, 15, Article 310. <https://doi.org/10.1186/s12913-015-0953-6>
- Davidson, M. (2014). Known-groups validity. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research*. Springer. https://doi.org/10.1007/978-94-007-0753-5_1581
- DeVellis, R. F. (2012). *Scale development: Theory and applications*. SAGE.
- Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: The development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science*, 9, Article 157. <https://doi.org/10.1186/s13012-014-0157-1>
- Ehrhart, M. G., Schneider, B., & Macey, W. H. (2014). *Organizational climate and culture: An introduction to theory, research, and practice*. Routledge.
- Gershon, R. R., Stone, P. W., Bakken, S., & Larson, E. (2004). Measurement of organizational culture and climate in healthcare. *Journal of Nursing Administration*, 34(1), 33–40. <https://doi.org/10.1097/00005110-200401000-00008>
- Ginty, A. T. (2013). Psychometric properties. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of behavioral medicine*. Springer. pp. 1563–1564. https://doi.org/10.1007/978-1-4419-1005-9_480
- Glisson, C., Green, P., & Williams, N. J. (2012). Assessing the organizational social context (OSC) of child welfare systems: Implications for research and practice. *Child Abuse and Neglect*, 36, 621–632.
- Glisson, C., & Hemmelgarn, A. (1998). The effects of organizational climate and interorganizational coordination on the quality and outcomes of children's service systems. *Child Abuse & Neglect*, 22(5), 401–421. [https://doi.org/10.1016/S0145-2134\(98\)00005-2](https://doi.org/10.1016/S0145-2134(98)00005-2)
- Glisson, C., Landsverk, J., Schoenwald, S., Kelleher, K., Hoagwood, K. E., Mayberg, S., & Green, P. (2008). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 35(1–2), 98–113. <https://doi.org/10.1007/s10488-007-0148-5>
- Glisson, C., & Williams, N. J. (2015). Assessing and changing organizational social contexts for effective mental health services. *Annual Review of Public Health*, 36, 507–523. <https://doi.org/10.1146/annurev-publhealth-031914-122435>
- Glisson, C., Williams, N. J., Green, P., Hemmelgarn, A., & Hoagwood, K. (2014). The organizational social context of mental health Medicaid waiver programs with family support services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 41, 32–42. <https://doi.org/10.1007/s10488-013-0517-1>
- Haynes, K., Nelson, K., & Blaine, D. D. (1999). Psychometric issues in assessment research. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 125–154). John Wiley & Sons.
- Helfrich, C. D., Li, Y.-F., Sharp, N. D., & Sales, A. E. (2009). Organizational readiness to change assessment (ORCA):

- Development of an instrument based on the promoting action on research in health services (PARIHS) framework. *Implementation Science*, 4, Article 38. <https://doi.org/10.1186/1748-5908-4-38>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. (2021). *Cochrane handbook for systematic reviews of interventions*. <https://training.cochrane.org/handbook/current/chapter-04>
- Jacobs, S. R., Weiner, B. J., & Bunger, A. C. (2014). Context matters: Measuring implementation climate among individuals and groups. *Implementation Science*, 9, Article 46. <https://doi.org/10.1186/1748-5908-9-46>
- Jung, T., Scott, T., Davies, H. T. W., Bower, P., & Whalley, D. (2009). Instruments for exploring organizational culture: A review of the literature. *Public Administration Review*, 69(6), 1087–1096. <https://doi.org/10.1111/j.1540-6210.2009.02066.x>
- Kimberly, J. R., & Cook, J. M. (2008). Organizational measurement and the implementation of innovations in mental health services. *Administration and Policy in Mental Health and Mental Health Services Research*, 35, 11–20. <https://doi.org/10.1007/s10488-007-0143-x>
- King, T. & Byers, J. F. (2007). A review of organizational culture instruments for nurse executives. *JONA: The Journal of Nursing Administration*, 37(1), 21–31. <https://doi.org/10.1097/00005110-200701000-00005>
- Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of Management Review*, 21(4), 1055–1080. <https://doi.org/10.5465/AMR.1996.9704071863>
- Lehman, W. E. K., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, 22, 197–209.
- Lewis, C. C., Boyd, M. R., Walsh-Bailey, C., Lyon, A. R., Beidas, R., Mittman, B., Aarons, G. A., Weiner, B. J., & Chambers, D. A. (2020). A systematic review of empirical studies examining mechanisms of implementation in health. *Implementation Science*, 15, Article 21.
- 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., Klasnja, P., Powell, B. J., Lyon, A. R., Tuzzio, L., Jones, S., & Walsh-Bailey, C. (2018). From classification to causality: Advancing Understanding of Mechanisms of change in implementation science. *Frontiers in Public Health*, 6, Article 136. <https://doi.org/10.3389/fpubh.2018.00136>
- Lewis, C. C., Mettert, K. D., Dorsey, C. N., Martinez, R. G., Weiner, B. J., Nolen, E., Stanick, C., 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>
- Lin, W.-L., & Yao, G. (2014). Concurrent validity. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research*. Springer. pp. 1184–1185. https://doi.org/10.1007/978-94-007-0753-5_516
- Mahoney, J. S., Ellis, T. E., Garland, G., Palyo, N., & Greene, P. K. (2012). Supporting a psychiatric hospital culture of safety. *Journal of the American Psychiatric Nurses Association*, 18(5), 299–306. <https://doi.org/10.1177/1078390312460577>
- 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>
- 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>
- Moise, I. K., Green, D., Toth, J., & Mulhall, P. F. (2014). Evaluation of an authority innovation-decision: Brief alcohol intervention for pregnant women receiving women, infants, and children services at two Illinois health departments. *Substance Use & Misuse*, 49(7), 804–812. <https://doi.org/10.3109/10826084.2014.880484>
- 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>
- Moos, R. H., Finney, J. W., & Gamble, W. (1982). The process of recovery from alcoholism. II. Comparing spouses of alcoholic patients and matched community controls. *Journal of Studies on Alcohol*, 43(9), 888–909. <https://doi.org/10.15288/jsa.1982.43.888>
- Parry, K. W., & Proctor-Thomson, S. B. (2001). Testing the validity and reliability of the Organizational Description Questionnaire (ODQ). *International Journal of Organisational Behaviour*, 4(3), 111–124.
- Patterson, M. G., West, M. A., Shackleton, V. J., Dawson, J. F., Lawthom, R., Maitlis, S., Robinson, D. L., & Wallace, A. M. (2005). Validating the organizational climate measure: Links to managerial practices, productivity and innovation. *Journal of Organizational Behavior*, 26, 379–408. <https://doi.org/10.1002/job.312>
- Powell, B. J., Stanick, C. F., Halko, H. M., Dorsey, C. N., Weiner, B. J., Barwick, M. A., Damschroder, L. J., Wensing, M., Wolfenden, L., & Lewis, C. C. (2017). Toward criteria for pragmatic measurement in implementation research and practice: A stakeholder-driven approach using concept mapping. *Implementation Science*, 12, Article 118. <https://doi.org/10.1186/s13012-017-0649-x>
- Proctor, E. K., Hooley, C., Morse, A., McCrary, S., Kim, H., & Kohl, P. L. (2019). Intermediary/purveyor organizations for evidence-based interventions in the US child mental health: Characteristics and implementation strategies. *Implementation Science*, 14, Article 3. <https://doi.org/10.1186/s13012-018-0845-3>
- Proctor, E. K., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G. A., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Ramsey, A. T., van den Berk-Clark, C., & Patterson, D. A. (2015). Provider-agency fit in substance abuse treatment organizations: Implications for learning climate, morale, and evidence-based practice implementation. *BMC Research Notes*, 8, Article 194. <https://doi.org/10.1186/s13104-015-1110-3>
- Rubenstein, L. V., Danz, M. S., Crain, A. L., Glasgow, R. E., Whitebird, R. R., & Solberg, L. I. (2014). Assessing

- organizational readiness for depression care quality improvement: Relative commitment and implementation capability. *Implementation Science*, 9, Article 173. <https://doi.org/10.1186/s13012-014-0173-1>
- Schein, E. H. (2017). *Organizational culture and leadership* (5th ed.). Wiley.
- Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual Review of Psychology*, 64, 361–388.
- Schneider, B., González-Romá, V., Ostroff, C., & West, M. A. (2017). Organizational climate and culture: Reflections on the history of the constructs in the Journal of Applied Psychology. *Journal of Applied Psychology*, 102, 468–482. <https://doi.org/10.1037/apl0000090>
- Scott, T., Mannion, R., Davies, H., & Marshall, M. (2003). The quantitative measurement of organizational culture in health care: A review of the available instruments. *Health Services Research*, 38(3), 923–945. <https://doi.org/10.1111/1475-6773.00154>
- Scott, V. C., Kenworthy, T., Godly-Reynolds, E., Bastien, G., Scaccia, J., McMickens, C., Rachel, S., Cooper, S., Wrenn, G., & Wandersman, A. (2017). The Readiness for Integrated Care Questionnaire (RICQ): An instrument to assess readiness to integrate behavioral health and primary care. *American Journal of Orthopsychiatry*, 87(5), 520–530. <https://doi.org/10.1037/ort0000270>
- Simpson, D. D. (2002). A conceptual framework for transferring research to practice. *Journal of Substance Abuse Treatment*, 22, 171–182.
- Stanick, C. F., Halko, H. M., Dorsey, C. N., Weiner, B. J., Powell, B. J., Palinkas, L. A., & Lewis, C. C. (2018). Operationalizing the “pragmatic” measures construct using a stakeholder feedback and a multi-method approach. *BMC Health Services Research*, 18, Article 882. <https://doi.org/10.1186/s12913-018-3709-2>
- Stanick, C. F., Halko, H. M., Nolen, E. A., Powell, B. J., Dorsey, C. N., Mettert, K. D., Weiner, B. J., Barwick, M., Wolfenden, L., Damschroder, L. J., & Lewis, C. C. (2021). Pragmatic measures for implementation research: Development of the Psychometric and Pragmatic Evidence Rating Scale (PAPERS). *Translational Behavioral Medicine*, 11, 11–20. <https://doi.org/10.1093/tbm/ibz164>
- 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 Treatment*, 32(3), 225–238. <https://doi.org/10.1016/j.jsat.2007.01.002>
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4), 651–657. <https://doi.org/10.1007/s11136-011-9960-1>
- Verbeke, W., Volgering, M., & Hessels, M. (1998). Exploring the conceptual expansion within the field of organizational behavior: Organizational climate and organizational culture. *Journal of Management Studies*, 35(3), 303–329. <https://doi.org/10.1111/1467-6486.00095>
- Weiner, B. J., Belden, C. M., Bergmire, D. M., & Johnston, M. (2011). The meaning and measurement of implementation climate. *Implementation Science*, 6, Article 78. <https://doi.org/10.1186/1748-5908-6-78>
- Weiner, B. J., Mettert, K. D., Dorsey, C. N., Nolen, E. A., Stanick, C., Powell, B. J., & Lewis, C. C. (2020). Measuring readiness for implementation: A systematic review of measures’ psychometric and pragmatic properties. *Implementation Research and Practice*, 1, 1–29. <https://doi.org/10.1177/2633489520933896>
- Williams, N. J. (2016). Multilevel mechanisms of implementation strategies in mental health: Integrating theory, research, and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(5), 783–798. <https://doi.org/10.1007/s10488-015-0693-2>
- Williams, N. J., Ehrhart, M. G., Aarons, G. A., Marcus, S. C., & Beidas, R. S. (2018). Linking molar organizational climate and strategic implementation climate to clinicians’ use of evidence-based psychotherapy techniques: Cross-sectional and lagged analyses from a 2-year observational study. *Implementation Science*, 13, Article 85. <https://doi.org/10.1186/s13012-018-0781-2>
- Williams, N. J., Frank, H. E., Frederick, L., Beidas, R. S., Mandell, D. S., Aarons, G. A., Green, P., & Locke, J. (2019). Organizational culture and climate profiles: Relationships with fidelity to three evidence-based practices for autism in elementary schools. *Implementation Science*, 14, Article 15. <https://doi.org/10.1186/s13012-019-0863-9>
- 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, 269–283. <https://doi.org/10.1007/s10488-016-0742-5>
- Williams, N. J., Wolk, C. B., Becker-Haimes, E. M., & Beidas, R. S. (2020). Testing a theory of strategic implementation leadership, implementation climate, and clinicians’ use of evidence-based practice: A 5-year panel analysis. *Implementation Science*, 15, Article 10. <https://doi.org/10.1186/s13012-020-0970-7>
- Wilson, P. M., Sales, A., Wensing, M., Aarons, G. A., Flottorp, S., Glidewell, L., Hutchinson, A., Presseau, J., Rogers, A., Sevdalis, N., Squires, J., & Straus, S. (2017). Enhancing the reporting of implementation research. *Implementation Science*, 12, Article 13. <https://doi.org/10.1186/s13012-017-0546-3>

Appendix I

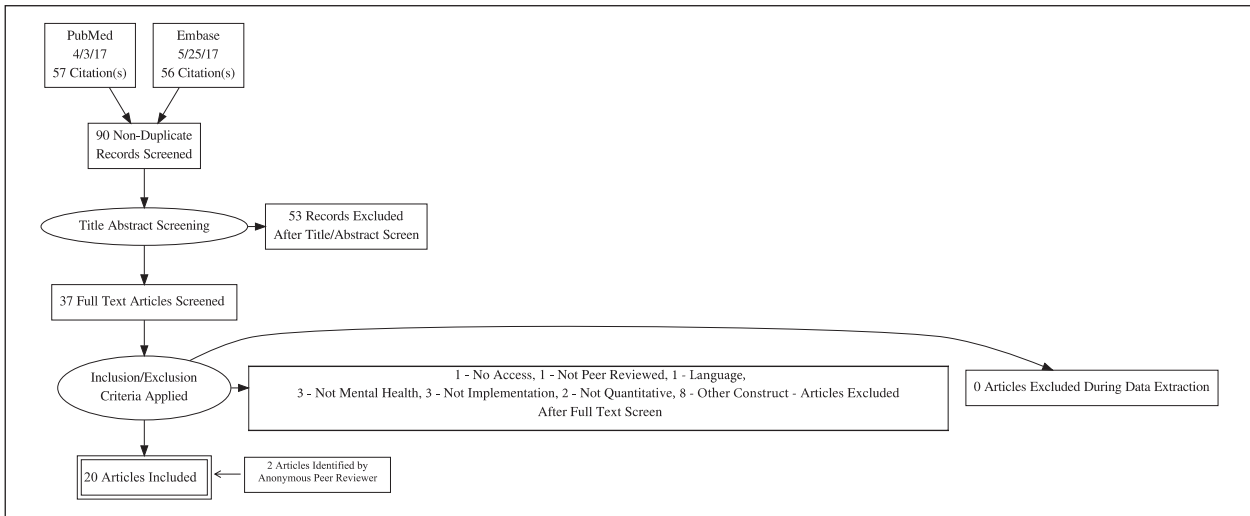


Figure 10. PRISMA diagram for organizational culture.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

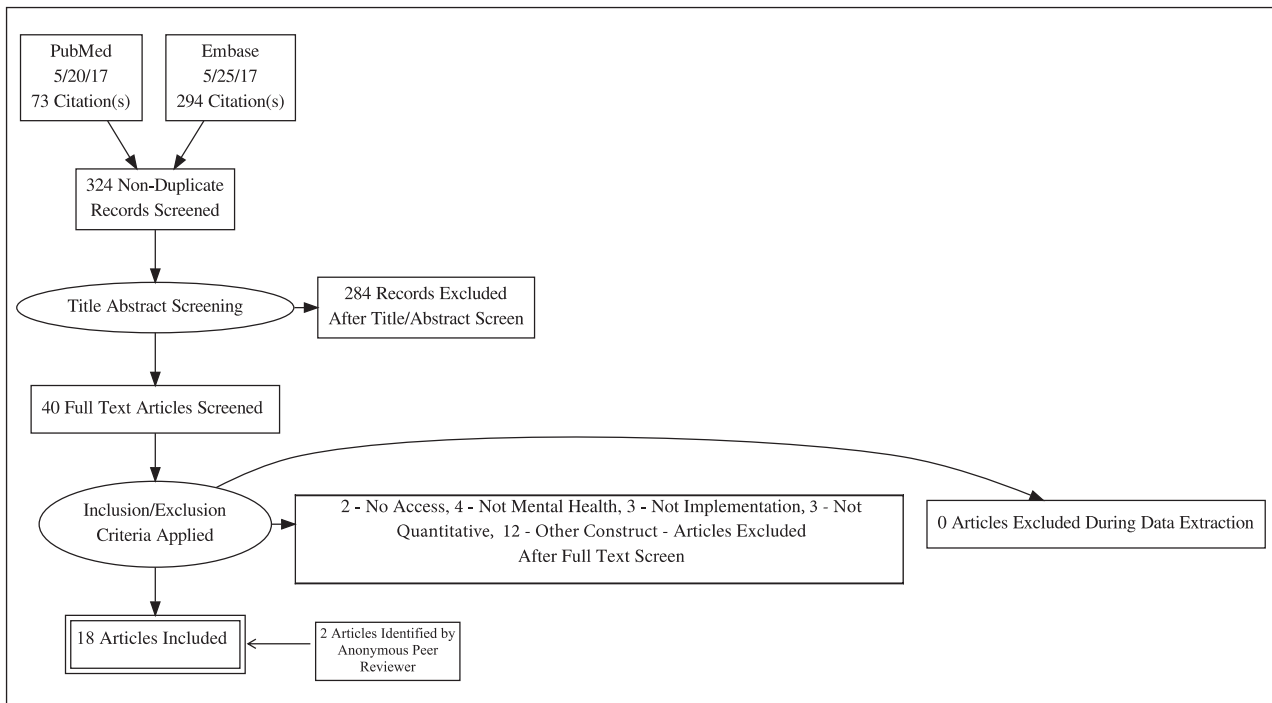


Figure 11. PRISMA diagram for implementation climate.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

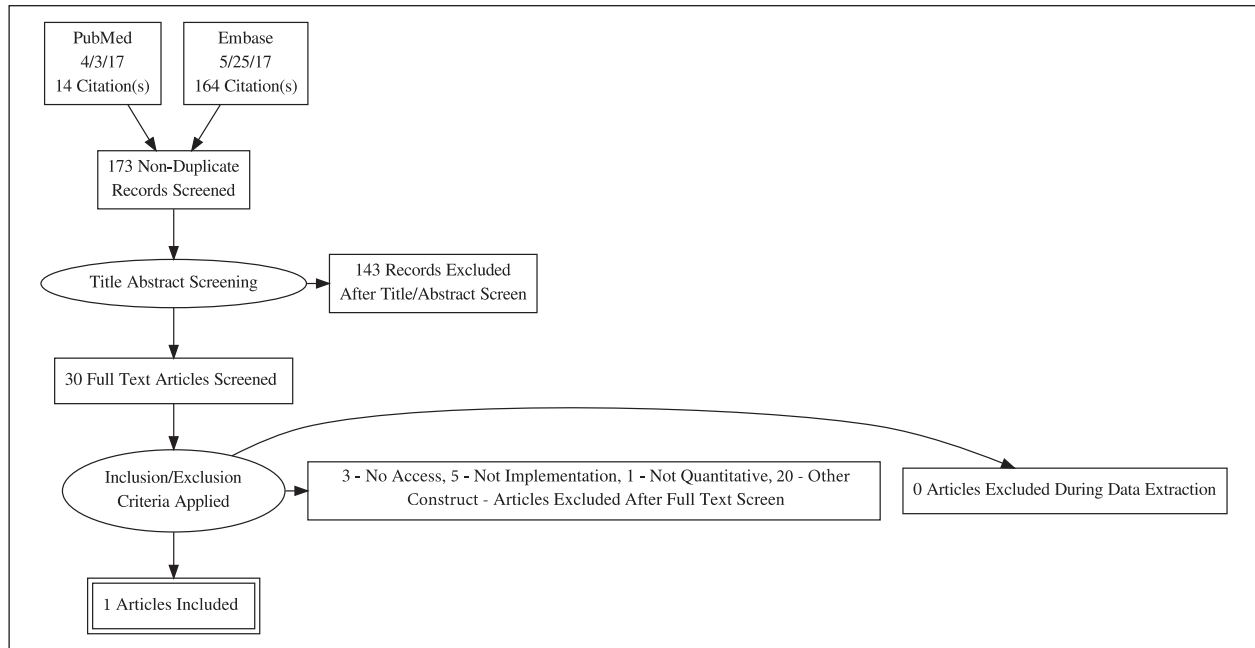


Figure 12. PRISMA diagram for tension for change.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

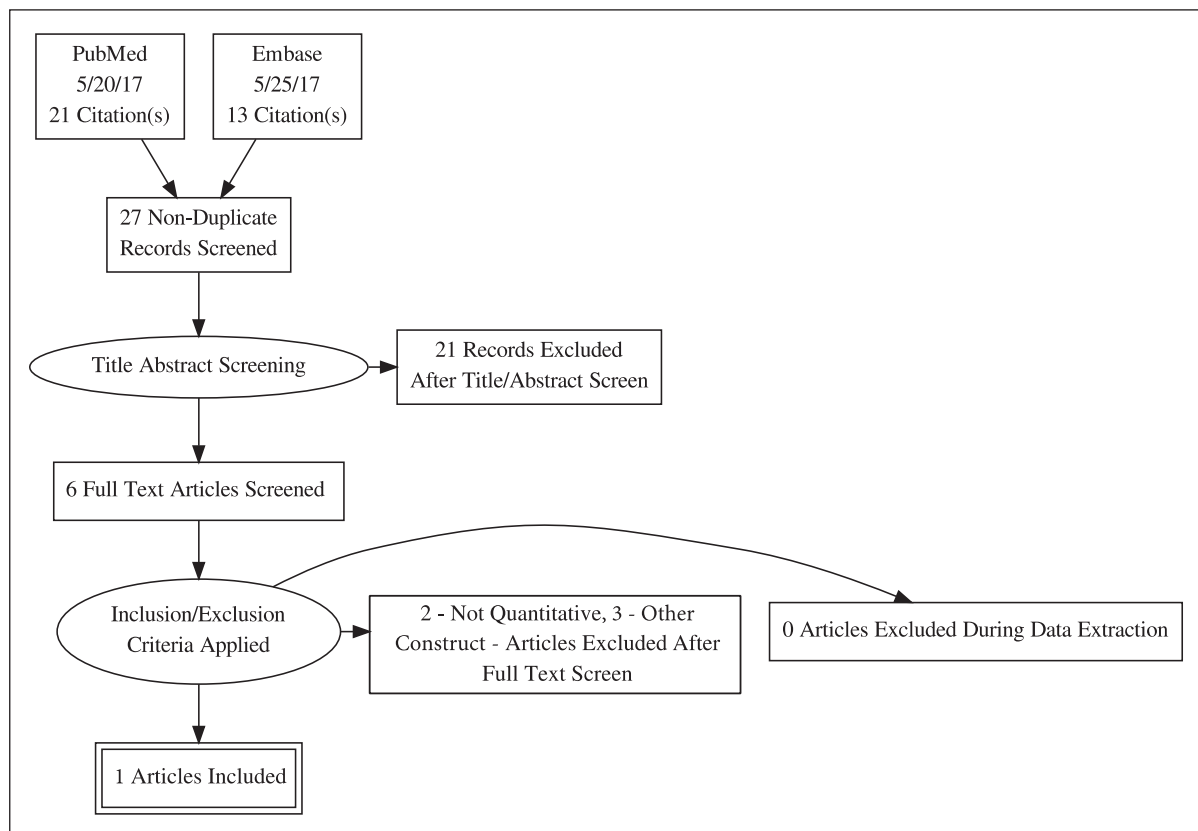


Figure 13. PRISMA diagram for compatibility.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

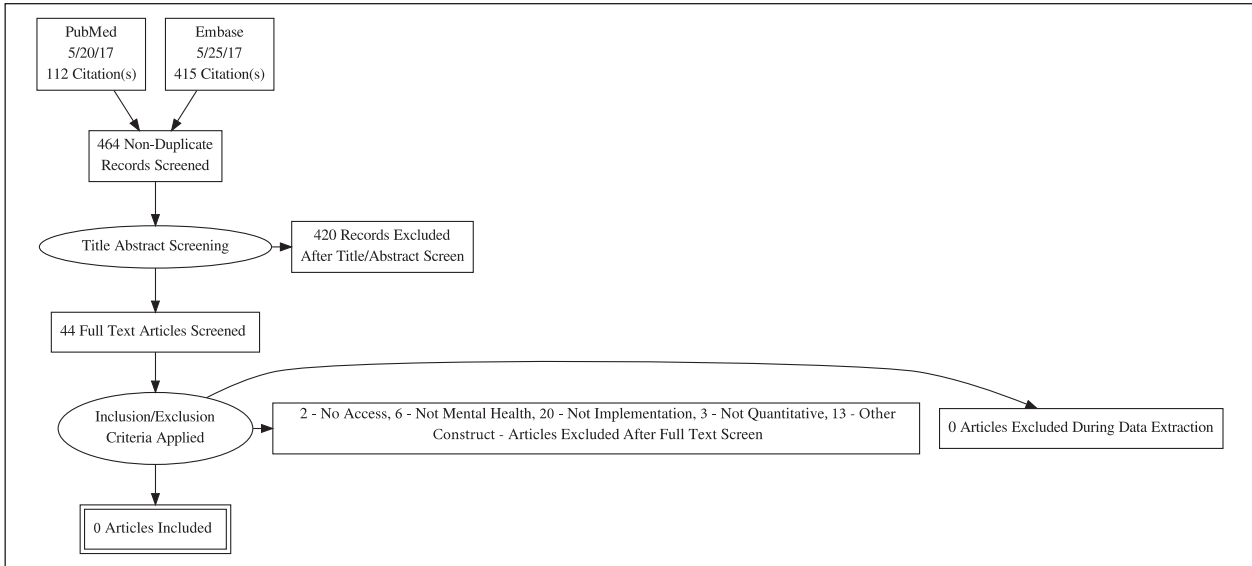


Figure 14. PRISMA diagram for relative priority.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

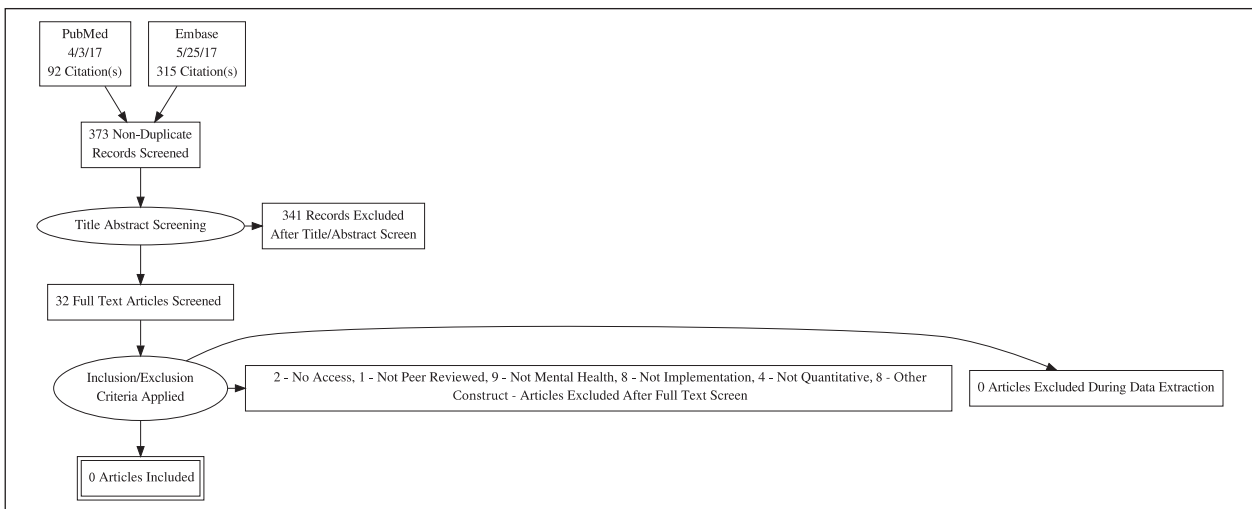


Figure 15. PRISMA diagram for organizational incentives and rewards.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

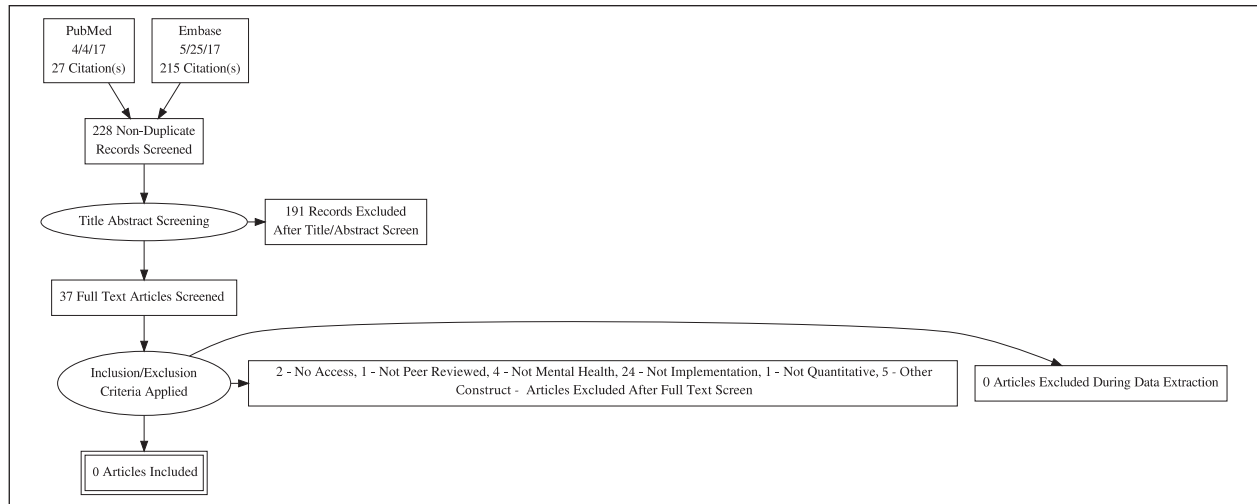


Figure 16. PRISMA diagram for goals and feedback.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.

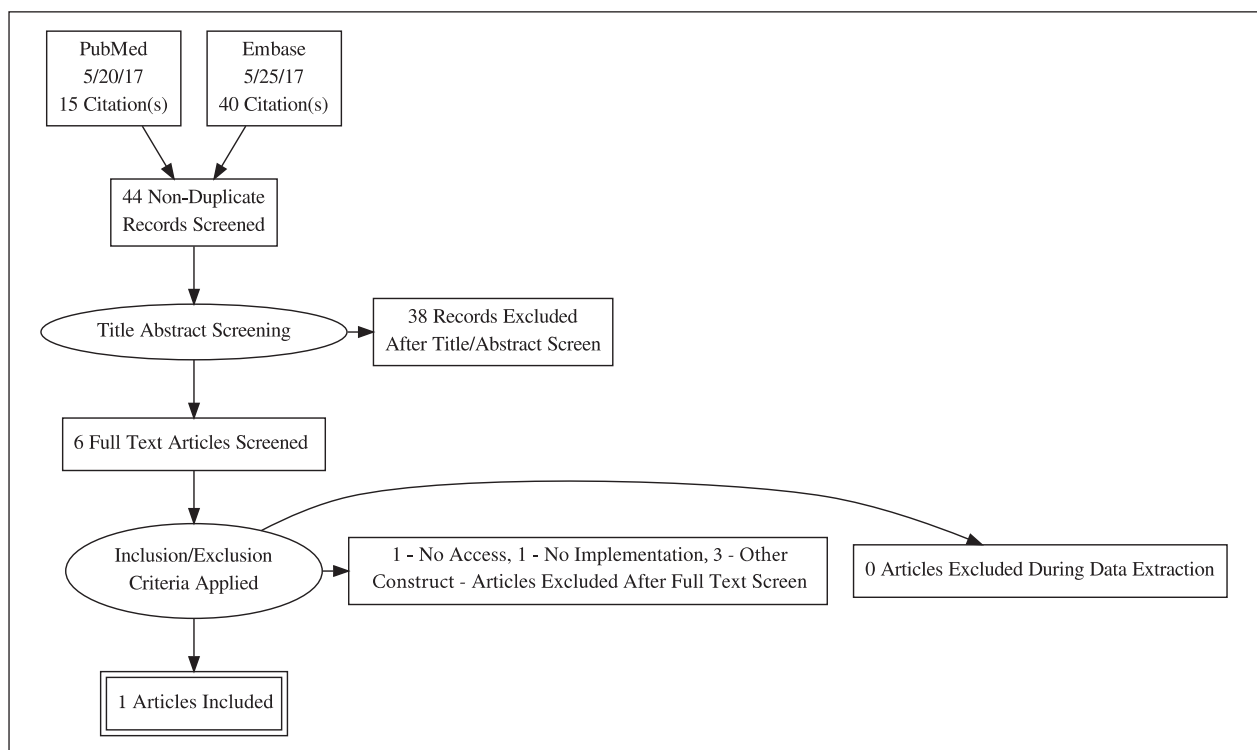


Figure 17. PRISMA diagram for learning climate.

Note that the number of articles identified did not equal the number of measures included in the analysis because in some cases a single measure was identified in multiple articles, and in others, multiple measures were identified in a single article.