



# A systematic review of measures of implementation players and processes: Summarizing the dearth of psychometric evidence

Implementation Research and Practice  
Volume 2: Jan-Dec 2021 1–27  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/26334895211002474  
journals.sagepub.com/home/irp  
 SAGE

Caitlin N Dorsey<sup>1</sup>, Kayne D Mettert<sup>1</sup> , Ajeng J Puspitasari<sup>2</sup>,  
Laura J Damschroder<sup>3</sup> and Cara C Lewis<sup>1,4</sup>

## Abstract

**Background:** Measurement is a critical component for any field. Systematic reviews are a way to locate measures and uncover gaps in current measurement practices. The present study identified measures used in behavioral health settings that assessed all constructs within the *Process* domain and two constructs from the *Inner setting* domain as defined by the Consolidated Framework for Implementation Research (CFIR). While previous conceptual work has established the importance social networks and key stakeholders play throughout the implementation process, measurement studies have not focused on investigating the quality of how these activities are being carried out.

**Methods:** The review occurred in three phases: Phase I, data collection included (1) search string generation, (2) title and abstract screening, (3) full text review, (4) mapping to CFIR-constructs, and (5) “cited-by” searches. Phase II, data extraction, consisted of coding information relevant to the nine psychometric properties included in the Psychometric And Pragmatic Rating Scale (PAPERS). In Phase III, data analysis was completed.

**Results:** Measures were identified in only seven constructs: *Structural characteristics* ( $n = 13$ ), *Networks and communication* ( $n = 29$ ), *Engaging* ( $n = 1$ ), *Opinion leaders* ( $n = 5$ ), *Champions* ( $n = 5$ ), *Planning* ( $n = 5$ ), and *Reflecting and evaluating* ( $n = 5$ ). No quantitative assessment measures of *Formally appointed implementation leaders*, *External change agents*, or *Executing* were identified. Internal consistency and norms were reported on most often, whereas no studies reported on discriminant validity or responsiveness. Not one measure in the sample reported all nine psychometric properties evaluated by the PAPERS. Scores in the identified sample of measures ranged from “-2” to “10” out of a total of “36.”

**Conclusions:** Overall measures demonstrated minimal to adequate evidence and available psychometric information was limited. The majority were study specific, limiting their generalizability. Future work should focus on more rigorous measure development and testing of currently existing measures, while moving away from creating new, single use measures.

**Plain Language Summary:** How we measure the processes and players involved for implementing evidence-based interventions is crucial to understanding what factors are helping or hurting the intervention’s use in practice and how to take the intervention to scale. Unfortunately, measures of these factors—stakeholders, their networks and communication, and their implementation activities—have received little attention. This study sought to identify and evaluate the quality of these types of measures. Our review focused on collecting measures used for identifying influential staff members,

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

<sup>2</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA

<sup>3</sup>Veterans Affairs (VA) Center for Clinical Management Research, Ann Arbor Healthcare System, Ann Arbor, MI, USA

<sup>4</sup>Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

## Corresponding author:

Caitlin N Dorsey, Kaiser Permanente Washington Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA 98122, USA.  
Email: Caitlin.N.Dorsey@kp.org



known as opinion leaders and champions, and investigating how they plan, execute, engage, and evaluate the hard work of implementation. Upon identifying these measures, we collected all published information about their uses to evaluate the quality of their evidence with respect to their ability to produce consistent results across items within each use (i.e., reliable) and if they assess what they are intending to measure (i.e., valid). Our searches located over 40 measures deployed in behavioral health settings for evaluation. We observed a dearth of evidence for reliability and validity and when evidence existed the quality was low. These findings tell us that more measurement work is needed to better understand how to optimize players and processes for the purposes of successful implementation.

### Keywords

Implementation science, measure, reliability, validity, process, networks, communications, change agents, opinion leaders, structural characteristics

How we measure the progression of an implementation endeavor is vital to understanding (1) which interventions are appropriate for which settings, (2) what factors act as barriers or facilitators, (3) whether implementation is progressing as expected, and (4) if desired outcomes are achieved. After all, “science is measurement” (Siegel, 1964) and our confidence in knowledge gained depends on how reliable (i.e., consistency across administrations) and valid (i.e., quality of inferences or claims) measures are (D. A. Cook & Beckman, 2006). To this end, valid measurement begins with a strong theoretical base (Nilsen, 2015). Given implementation is multidimensional and complex, use of a single or universal theory is argued to offer only partial understanding. Alternatively, determinant frameworks embrace many theories, present a systems approach, and outline what hypothesized or established factors (i.e., barriers and facilitators) influence implementation outcomes.

Measurement—a foundational step in developing a strong empirical infrastructure for any scientific field (D. A. Cook & Beckman, 2006; Martinez et al., 2014; Stichter & Conroy, 2004) is often slower to progress and at times considered secondary. Studies may “economize” measurement in an attempt to reduce burden and/or generate more data; however, unknown measure reliability could lead to meaningful information not being collected (DeVellis, 2003). At least 10 (Chaudoir et al., 2013; Chor et al., 2015; Emmons et al., 2012; Hrisos et al., 2009; Ibrahim & Sidani, 2015; King & Byers, 2007; Lewis et al., 2015; T. Scott et al., 2003; Squires et al., 2011; Weiner et al., 2008) recent implementation science and quality improvement (QI) measurement-focused reviews reveal the same general conclusion: psychometric assessments are underreported and when available, measures often perform below acceptable standards. These reviews largely focus on the organization-, provider-, patient- and innovation-level (Chaudoir et al., 2013).

Measurement relies on definitions of fundamental concepts, encapsulated as constructs (i.e., a complex idea or concept formed by synthesizing simpler ideas; (American Psychological Association, 2020). The Consolidated

Framework for Implementation Research (CFIR; Damschroder et al., 2009) comprises 37 constructs believed to predict, moderate, or mediate implementation factors. The CFIR was developed based on a review of 500 published sources across 13 research areas (Greenhalgh et al., 2004) and 18 additional sources to centralize and unify this myriad of frameworks, create consistent language, and minimize overlap and redundancies. Constructs are organized into five domains for evaluation.

One domain influencing implementation effectiveness, *Process*, cuts across the other four CFIR domains, as it informs the approach or work processes for adapting the intervention to fit the setting (i.e., intervention characteristics), the context through which the implementation proceeds (i.e., inner and outer setting), and the actions and behaviors of the individuals involved who actively promote implementation (i.e., characteristics of individuals). The *Process* domain includes *Planning*, *Engaging*, *Reflecting and evaluating*, and *Executing* constructs. The CFIR further delineates *Engaging* to recognize the importance of involving the appropriate individuals in the implementation: *Opinion leaders*, *Formally appointed internal implementation leaders*, *Champions*, and *External change agents*. In the last decade since the CFIR was published, significant progress has underscored how these key actors influence the implementation process. Opinion leaders exert influence over the behaviors and/or attitudes of individuals within their social network with relative frequency (Rogers, 2003). Utilizing opinion leaders to promote and deliver new practices accelerates adoption (Valente, 2010). A review of 18 randomized controlled trials revealed health interventions as successfully promoted when implemented by opinion leaders alone or in combination with other strategies (e.g., audit and feedback, trainings, toolkits) (Flodgren et al., 2019). Champions are active in supporting and inspiring commitment to implementation, along with combatting indifference and resistance evidence-based practices may provoke (Damschroder et al., 2009; Howell & Higgins, 1990; Rogers, 2003). In substance use and mental health settings ( $n=13$  studies), a systematic review found clinical champions important for

facilitating practice change, overcoming barriers, and enhancing staff engagement (Wood et al., 2020). Champions and other internal implementation leaders are formally or informally appointed and can leverage their organizational position to facilitate the implementation process (Soo et al., 2009). An “informal” champion, on their own initiative, learns about an innovation (e.g., attending a conference session) returning enthusiastic to endorse its value for improving care. Organization leaders already demonstrating champion-like behaviors are formally appointed as executive champions (i.e., senior leadership) or managerial champions (i.e., clinical department, ward, or unit managers). Individuals outside the organization (i.e., external change agents) are also formally brought on board to facilitate or influence implementation. External change agents are professionally trained in organizational change science and/or the intervention being introduced (Damschroder et al., 2009). These *Process* constructs provide a lens for empirically testing the importance of these actor influences. To our knowledge, no research to date has assessed the quality of measures relevant to *Process* constructs. Implicit in these studies is that these actors can leverage their social and professional networks (*Networks and communications* within CFIR’s *Inner setting* domain) to promote communication and collaboration among intervention recipients and other implementation stakeholders (e.g., researchers, community members, policy makers, and practitioners) (Rogers, 2003). These connections are key, as a lack of communication between stakeholders is one of the most commonly occurring barriers to successful implementation (Rycroft-Malone et al., 2013). Using the CFIR to guide their evaluations, both MOVE! (Damschroder & Lowery, 2013) and Telephone Lifestyle Coaching (TLC; Damschroder et al., 2017) programs revealed the *Networks and communications* construct to have strong, positive impacts on implementation outcomes. These findings demonstrate the importance of strong working relationships and communications between implementation leaders (MOVE!) or coordinator and physician champions (TLC). Furthermore, *Structural characteristics* (within CFIR’s *Inner setting* domain) defines how an organization’s social architecture impacts stakeholder communication. Illustrated by a recent implementation of an opioid-use treatment initiative across multiple settings (e.g., established systems with well-informed, invested, long-standing staff) and levels (e.g., newer provider-level structures including staff with a high degree of excitement and enthusiasm) communication was identified as a critical component influencing success (Hanna et al., 2020).

This systematic review sought to identify measures evaluating constructs outlined in the *Process* domain of the CFIR (Damschroder et al., 2009): (1) *Planning*, (2) *Engaging*, (3) *Opinion leaders*, (4) *Formally appointed internal implementation leaders*, (5) *Champions*, (6) *External change agents*, (7) *Executing*, and (8) *Reflecting and evaluating*. We also investigated two constructs from

the *Inner setting* domain (9) *Structural characteristics* and (10) *Networks and communications* (Table 1) for the reasons above. Consistent with our funding source (MH106510), this review focused on identifying measures used in behavioral health settings. Empirical evidence for identified measures was assessed and rated for quality using a portion of the Psychometric and Pragmatic Evaluation Rating Scale (PAPERS) (Stanick et al., 2021).

## Methods

### Design overview

Data were collected as part of a larger initiative spearheaded by several Society for Implementation Research Collaboration (SIRC) investigators aiming to identify reliable and valid quantitative implementation-related measures used in behavioral health. The current study is one of a series of systematic reviews that identify and evaluate measures assessing constructs for both the CFIR and Implementation Outcomes Framework. Full details of the protocol for the entire set of systematic reviews have been published elsewhere (Lewis et al., 2018).

Methods included three phases. Phase I, data collection, included five steps: a) search string generation, b) title and abstract screening, c) full text review, d) mapping to CFIR-constructs, and e) “cited-by” searches. Phase II, data extraction, coded relevant psychometric information, and in Phase III data analysis was completed.

### Phase I: data collection

Literature searches were conducted in PubMed and Embase bibliographic databases using search strings curated in consultation with PubMed support specialists and a library scientist. Consistent with our aim to identify and assess implementation-related measures in behavioral health, searches included four core levels: 1) implementation terms (e.g., diffusion, knowledge translation, adoption); 2) measurement terms (e.g., instrument, survey, questionnaire); 3) evidence-based practice terms (e.g., innovation, guideline, empirically supported treatment); and 4) behavioral health terms (e.g., behavioral medicine, mental disease, psychiatry) (BLINDED CITATION). A fifth level of terms included each CFIR-construct: (1) *Structural characteristics*, (2) *Networks and communication*, (3) *Planning*, (4) *Engaging*, (5) *Opinion leaders*, (6) *Formally appointed internal implementation leaders*, (7) *Champions*, (8) *External change agents*, (9) *Executing*, and (10) *Reflecting and evaluating*. Articles published from 1985 on and in English were included. Literature searches were conducted independently for each construct from April–May 2017 (Table 2 lists construct-specific searches). Identified articles were vetted through a title and abstract screening, followed by full text review to confirm relevance. In brief, we included empirical studies

using quantitative measures to evaluate a behavioral health implementation effort.

The construct mapping phase assigned measures to any of the ten CFIR-constructs based on the author's description or dual-construct review (Lewis et al., 2018). Measures often include multiple scales, which are collections of items that are combined and scored to reveal levels of the concept being investigated; both full measures and individual scales were mapped. Content experts confirmed construct mappings through item-level assessments for each measure and individual scales part of separate, multi-scaled measures, if at least two-items were deemed relevant. In addition, searches from other constructs evaluated by the larger initiative were assessed for relevant measures. The final step encompassed "cited-by" searches in PubMed and Embase to identify empirical articles published from 1985 to 2019 that used the measure in behavioral health implementation research.

### Phase II: data extraction

Once all relevant literature was retrieved, articles were compiled into "measure packets," that included the measure itself (as available), article(s) describing its development (or the first empirical use in a behavioral health context), and all additional empirical uses in behavioral health. Authors reviewed each article and electronically coded all relevant reports of psychometric information pertaining to criteria included the Psychometric and Pragmatic Evaluation Rating Scale (PAPERS). Details on the full PAPERS rating system and criteria anchor descriptions is published elsewhere (Lewis et al., 2018). The nine psychometric properties in the PAPERS were used for rating: (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. Descriptive information was collected on: (1) whether the construct was defined; (2) number of uses; (3) number of items; (4) country of origin; (5) setting administered; (6) level of analysis; (7) population assessed; (8) stage of implementation defined by the Exploration, Adoption/Preparation, Implementation, Sustainment (EPIS) model (Aarons et al., 2011); and (9) if any implementation outcomes were assessed (i.e., acceptability, appropriateness, adoption, cost, feasibility, fidelity, penetration, sustainability; (Proctor et al., 2011). Data were collected at both the full measure and individual scale level. The setting administered, level of analysis, and population assessed categories used an exhaustive list of potential codes, developed internally based on our team's content expertise, to capture the most comprehensive data set possible.

Using coded information from each article, measure packets were then rated by the PAPERS. Each criterion was rated on the following scale: "poor" (-1), "none" (0),

"minimal/emerging" (1), "adequate" (2), "good" (3), or "excellent" (4). The "poor" (-1) rating was used when psychometric testing was conducted and demonstrated results inconsistent with study hypotheses, while "none" (0) indicated no tests were conducted, therefore no evidence present. Two raters independently applied the PAPERS and came to consensus. Total scores were calculated by summing each individual criterion score for a maximum possible total score of 36 and a minimum possible total score of -9. Higher PAPERS scores indicated higher psychometric quality.

For measure packets with multiple articles, the median rating across all articles was calculated for each criterion and then summed. In cases where the median of two scores would equal "0" (e.g., a score of -1 and 1), the lower score was retained. If no psychometric information was available on the full measure level, but multiple individual scales were relevant to the construct, we "rolled-up" all available scale scores. The roll-up median approach was adopted to offer a less negative assessment of each property by considering all empirical uses instead of the commonly used "worst score counts" approach (Lewis et al., 2015; Terwee et al., 2012) where only the one lowest performing article was considered in cases of multiple ratings. Once the roll-up process was complete, median ratings were calculated as needed for each criterion and summed to determine a total score.

### Phase III: data analysis

Simple statistics (i.e., frequencies) were calculated to report on the presence and quality of psychometric-relevant information. Each measure was assigned a total quality score encapsulating all nine psychometric criteria scores. Bar charts were generated to display visual head-to-head comparisons within a given construct.

## Results

### Measure search results

Searches from PubMed and Embase electronic databases revealed a total of 45 unique measures ( $n=24$  full measures and  $n=21$  individual scales). Please see Figures A1-A10, which can be found in the Supplemental Appendix, for PRISMA flowcharts of included and excluded studies. At least one measure was identified for 7 of the 10 constructs. No quantitative assessment measures of *Formally appointed implementation leaders*, *External change agents*, or *Executing* were identified. Each unique measure had the potential of being mapped to more than one construct: *Structural characteristics* ( $n=13$ ), *Networks and communication* ( $n=29$ ), *Engaging* ( $n=1$ ), *Opinion leaders* ( $n=5$ ), *Champions* ( $n=5$ ), *Planning* ( $n=5$ ), and *Reflecting and evaluating* ( $n=5$ ). In addition, we mapped



**Table 1.** Definitions of process and inner setting constructs.

General process	Successful implementation usually requires an active change process aimed to achieve individual and organizational level use of the intervention as designed. Individuals may actively promote the implementation process and may come from the inner or outer setting (e.g., local champions, external change agents).
Planning	The degree to which a scheme or method of behavior and tasks for implementing an intervention are developed in advance and the quality of those schemes or methods
Engaging	Attracting and involving appropriate individuals in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, and other similar activities
Opinion leaders	Individuals in an organization who have formal or informal influence on the attitudes and beliefs of their colleagues with respect to implementing the intervention
Formally appointed internal implementation leaders	Individuals from within the organization who have been formally appointed with responsibility for implementing an intervention as coordinator, project manager, team leader, or other similar role
Champions	Individuals who dedicate themselves to supporting, marketing, and 'driving through' an implementation, overcoming indifference or resistance that the intervention may provoke in an organization
External change agents	Individuals who are affiliated with an outside entity who formally influence or facilitate intervention decisions in a desirable direction
Executing	Carrying out or accomplishing the implementation according to plan
Reflecting and Evaluating	Quantitative and qualitative feedback about the progress and quality of implementation accompanied with regular personal and team debriefing about progress and experience
Structural characteristics	The social architecture, age, maturity, and size of an organization
Networks and communications	The nature and quality of webs of social networks and the nature and quality of formal and informal communications within an organization

Source: From Damschroder et al. (2009).

to the “*General process*” domain ( $n=4$ ) when measures could not be assigned to the nuanced constructs, but clear relevance to the *Process* domain was observed.

### Measure characteristics

Most measures were developed in the United States ( $n=34$ ; 75.56%), with other countries of origin including Australia, Canada, Netherlands, Norway, Spain, and the United Kingdom. Measures were most frequently administered in outpatient community settings ( $n=24$ ; 53.33%) and during implementation efforts targeting general mental health problems ( $n=19$ ; 42.22%) or substance-use disorders ( $n=17$ ; 37.78%). The unit of analysis was typically at the provider-level ( $n=31$ ; 68.89%), as well as at the clinic- or site- ( $n=11$ ; 24.44%), director- ( $n=11$ ; 24.44%), and supervisor-levels ( $n=9$ ; 20%). Results were mixed when examining the number of uses: *Structural characteristics* ( $n=10$ ) had eight (80%) measures that were used only once, whereas *Networks and communications* ( $n=24$ ) had 13 (54%) measures that were used more than once. Measures assessed only three of the eight implementation outcomes: Adoption (*Networks and communications*, *Opinion leaders*, *Champions*), Fidelity (*Networks and communications* and *General process*), and Sustainability (*General process*). Please see Table 3 for a breakdown of all measure characteristics by each construct and the *General Process* domain-level category.

### Psychometric and Pragmatic Evidence Rating Scale (PAPERS) results

Of the 45 unique measures, close to one-third ( $n=14$ ; 31.11%) could not be rated using the PAPERS. That is, these measures could not be evaluated for psychometric quality from a classical test theory perspective, for which PAPERS is designed. For example, to identify coworkers as opinion leaders, counselors were asked two questions: (1) *From among your coworkers, whom would you go to if you had questions about general treatment issues for a client?* and (2) *From among your coworkers, whom would you go to if you had questions about a client who might have co-occurring mental health and substance abuse problems?* and a network analysis was conducted to establish links between the nominated opinion leaders and their peers (Moore et al., 2004). These measures were kept in our sample as they still employed quantitative methods.

Overall, psychometric information was limited for the remaining 31 measures (Table 4). There were no measures containing information on all nine PAPERS criteria. No reports of discriminant validity or responsiveness were captured. Internal consistency and norms were reported in all constructs except for *Engaging*. Bar charts displaying visual head-to-head comparisons across constructs were generated (Figures 1 to 6). Total PAPERS scores ranged from -2 to 10 and generally were “1—minimal/emergent” to “2—adequate,” with higher scores indicating higher

psychometric quality. Table 5 includes PAPERS summary median scores and ranges for measures that could be rated ( $n=31$ ; 68.89%). The Organizational Readiness for Change (ORC) Communication scale (Lehman et al., 2002) reported the most psychometric properties: internal consistency (“3—good”), convergent validity (“2—adequate”), known-groups validity (“2—adequate”), predictive validity (“1—minimal/emerging”), and norms (“2—adequate”). The ORC Communication scale, along with the Implementation Phases Inventory (Bradshaw et al., 2009) scored the highest on the PAPERS, with an overall rating of 10 out of a possible 36.

*Structural characteristics* ( $N=13$  measures and scales). Four (30.77%) measures could not be evaluated using the PAPERS (Beidas et al., 2015; Lundgren et al., 2011; Ramsay et al., 2016; Schoenwald et al., 2008). For example, in one study, supervisors provided research staff with the number of therapists in their unit and their employment status to determine program size and the percentage of fee-for-service staff (Beidas et al., 2015). Six measures contained psychometric data and reported internal consistency ( $n=5$ ; 50%), predictive validity ( $n=1$ ; 10%), and norms ( $n=5$ ; 50%). Total PAPERS scores ranged from -2 to 6. Psychometric ratings for each *Structural characteristics* measure are described in Table 6. Only individual scales from multi-scaled measures had psychometric information available. In the Nieboer Formulation Scale (Nieboer & Strating, 2012), two of the four individual scales were relevant and produced different internal consistency scores: Job specificity ( $\alpha=0.65$ ; “1—minimal/emerging”) and Task routineness ( $\alpha=0.82$ ; “3—good”).

*Networks and communications* ( $N=29$  measures and scales). Five (17.24%) measures were not suitable for rating (Feinberg et al., 2005; Hanbury, 2013; Johnson et al., 2015; Li et al., 2012; Palinkas et al., 2011). One such study implemented an evidence-based treatment for externalizing behaviors and mental health problems, asking administrators to identify individuals they relied on for advice to determine influential network actors (Palinkas et al., 2011). In the remaining 24 measures, seven of the nine PAPERS criteria were reported: internal

consistency ( $n=17$ ; 70.83%), convergent validity ( $n=2$ ; 8.33%), known-groups validity ( $n=2$ ; 8.33%), predictive validity ( $n=6$ ; 25%), concurrent validity ( $n=2$ ; 8.33%), structural validity ( $n=1$ ; 4.17%), and norms ( $n=20$ ; 83.33%). Total PAPERS scores ranged from -1 to 11. Median ratings for internal consistency and structural validity were “3—good,” “2—adequate” for convergent validity, known-groups validity, and norms, and “1—minimal/emerging” for predictive validity and concurrent validity. The only measure in the entire study sample to report on structural validity, The Team Climate Inventory (Anderson & West, 1998), mapped to this construct, receiving a “3—good.” Please refer to Table 7 for psychometric ratings for each *Networks and communications* measure.

*General process* ( $n=4$  measures and scales). This assignment was used when measures did not fit one specific CFIR construct but were still relevant to the *Process* domain. Three measure contained psychometric information: internal consistency ( $n=2$ ; 50%), convergent validity ( $n=1$ ; 25%), known-groups validity ( $n=1$ ; 25%), predictive validity ( $n=2$ ; 50%), concurrent validity ( $n=2$ ; 50%), and norms ( $n=3$ ; 75%). Total PAPERS scores ranged from 2 to 10, with median ratings including internal consistency (“2—adequate”), convergent validity (“2—adequate”), known-groups validity (“4—excellent”), predictive validity (“-1—poor”), concurrent validity (“2—adequate”), and norms (“3—good”). Table 8 includes a breakdown of each individual property rating for *General process* measures. The Implementation Phases Inventory (Bradshaw et al., 2009), one of the top performing measures, mapped to *General Process* and scored a 10 out of 36 on the PAPERS.

*Planning* ( $n=5$  scales). All individual scales reported psychometric information, with total PAPERS scores ranging from 1 to 7. Only two criterion had information available: internal consistency ( $n=4$ ; 80%) and norms ( $n=3$ ; 60%). Table 9 contains each *Planning* scale’s ratings. Internal consistency had a “3—good” median score and a “1—minimal/emerging” median score for norms. The best performing *Planning* scale contained 15 items and was used in three behavioral health relevant studies. This individual scale is one of four scales included in the

**Table 2.** Electronic bibliographic database search terms.

Pubmed electronic database	
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

(Continued)

**Table 2.** (Continued)

Pubmed electronic database

Search term	Search string
Measurement	(instrument[tw] OR (survey[tw] OR survey'[tw] OR survey's[tw] OR survey100[tw] OR survey12[tw] OR survey1988[tw] OR survey226[tw] OR survey36[tw] OR surveyability[tw] OR surveyable[tw] OR surveyance[tw] OR surveyans[tw] OR surveyansin[tw] OR surveybetween[tw] OR surveyd[tw] OR surveydagger[tw] OR surveydata[tw] OR surveydelhi[tw] OR surveyed[tw] OR surveyedandtestedthe[tw] OR surveyedpopulation[tw] OR surveyees[tw] OR surveyelicited[tw] OR surveyer[tw] OR surveyes[tw] OR surveyeyed[tw] OR surveyform[tw] OR surveyfreq[tw] OR surveygizmo[tw] OR surveyin[tw] OR surveying[tw] OR surveying' [tw] OR surveyings[tw] OR surveylogistic[tw] OR surveymaster[tw] OR surveymeans[tw] OR surveymeter[tw] OR surveymonkey[tw] OR surveymonkey's[tw] OR surveymonkeytrade[tw] OR surveyng[tw] OR surveyor[tw] OR surveyor' [tw] OR surveyor's[tw] OR surveyors[tw] OR surveyors' [tw] OR surveyortrade[tw] OR surveypatients[tw] OR surveyphreg[tw] OR surveyplus[tw] OR surveyprocess[tw] OR surveyreg[tw] OR surveys[tw] OR surveys' [tw] OR surveys'food[tw] OR surveys'usefulness[tw] OR surveysclub[tw] OR surveyselect[tw] OR surveyset[tw] OR surveyset' [tw] OR surveyspot[tw] OR surveystrade[tw] OR surveysuite[tw] OR surveytaken[tw] OR surveythese[tw] OR surveytm[tw] OR surveytracker[tw] OR surveytrade[tw] OR surveyvas[tw] OR surveywas[tw] OR surveywiz[tw] OR surveyxact[tw]) OR (questionnaire[tw] OR questionnaire' [tw] OR questionnaire'07[tw] OR questionnaire'midwife[tw] OR questionnaire's[tw] OR questionnaire1[tw] OR questionnaire11[tw] OR questionnaire12[tw] OR questionnaire2[tw] OR questionnaire25[tw] OR questionnaire3[tw] OR questionnaire30[tw] OR questionnaireand[tw] OR questionnairebased[tw] OR questionnairebefore[tw] OR questionnaireconsisted[tw] OR questionnairecopyright[tw] OR questionnaired[tw] OR questionnairedeveloped[tw] OR questionnaireepq[tw] OR questionnaireforpediatric[tw] OR questionnairegrt[tw] OR questionnairehas[tw] OR questionnaireitaq[tw] OR questionnairel02[tw] OR questionnairemcesqscale[tw] OR questionnaireurse[tw] OR questionnaireon[tw] OR questionnaireonline[tw] OR questionnairepf[tw] OR questionnairephq[tw] OR questionnaires[tw] OR questionnaires' [tw] OR questionnaires''[tw] OR questionnairescan[tw] OR questionnairesdq1 adolescent[tw] OR questionnairess[tw] OR questionnairestrade[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 measurement1[tiab] OR measuremental[tiab] OR measurementd[tiab] OR measuremented[tiab] OR measurementexhaled[tiab] OR measurementf[tiab] OR measurementin[tiab] OR measuremention[tiab] OR measurementis[tiab] OR measurementkcomputation[tiab] OR measurementl[tiab] OR measurementmanometry[tiab] OR measurementmethods[tiab] OR measurementof[tiab] OR measurementon[tiab] OR measurementpro[tiab] OR measurementresults[tiab] OR measurements[tiab] OR measurements' [tiab] OR measurements's[tiab] OR measurements0[tiab] OR measurements5[tiab] OR measurementsa[tiab] OR measurementsare[tiab] OR measurementscanbe[tiab] OR measurementscheme[tiab] OR measurementsfor[tiab] OR measurementsgave[tiab] OR measurementsin[tiab] OR measurementsindicate[tiab] OR measurementsmoking[tiab] OR measurementsof[tiab] OR measurementson[tiab] OR measurementsreveal[tiab] OR measurementss[tiab] OR measurementswere[tiab] OR measurementtime[tiab] OR measurementts[tiab] OR measurementusing[tiab] OR measurementws[tiab]) OR measures[tiab] OR inventory[tiab]) 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 guideline1[tiab] OR guideline2015[tiab] OR guidelinebased[tiab] OR guidelineed[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 guidelines1[tiab] OR guidelines19[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 guidelinesthat[tiab] OR guidelinestrade[tiab] OR guidelineswiki[tiab]) OR "guidelines as topic"[MeSH Terms] OR "best practice*" [tw]) AND

(Continued)

**Table 2.** (Continued)

Pubmed electronic database	
Search term	Search string
Behavioral health	("mental health"[tw] OR "behavioral health"[tw] OR "behavioral health"[tw] OR q"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
Structural characteristics	"structural characteristic*" OR "social architecture" OR "age of organization" OR "maturity of organization" OR "size of organization" OR "functional differentiation" OR "administrative intensity" OR centralization[tw] OR
Networks and communication	networks [tw] OR communication [tw] OR communicate[tw] OR communicates[tw] OR "social networks" [tw] OR "formal communication"[tw] OR "informal communication"[tw] OR "internal bonding" OR teamness OR "shared vision" OR "information sharing"[tw] OR
Planning	planning [tw] OR health planning[mh] OR "comprehensive approach*" OR "implementation plan*" OR "tailored implementation*" OR strategies [tw] OR strategy[tw] OR workaround*[tw] OR formulate[tw] OR formulation[tw] OR formulated[tw] OR "course* of action"[tw] OR "implementation plans" OR "tailored implementation" OR strategies OR workarounds OR formulate OR "course* of action" OR
Engaging	engaged[tw] OR engages[tw] OR engagement[tw] OR attract*[tw] OR involv*[tw] OR "social market*" OR model[tw] OR models[tw] OR models, theoretical [mh] OR
Opinion leaders	"opinion leaders" OR
Formally appointed internal implementation leaders	"implementation leader*" [tw] OR coordinator[tw] OR "project manager*" [tw] OR "team leader*" [tw] OR
Champions	champion[tw] OR champions[tw] OR championed[tw] OR "transformational leader*" OR campaigner*[tw] OR promoter*[tw] OR proponent*[tw] OR supporter*[tw] OR
External change agents	"external change agent" OR consultant*[tw] OR "technical assist*" [tw] OR facilitator OR
Executing	execute[tw] OR executed[tw] OR execution[tw] OR "carr* out" [tw] OR accomplish*[tw] OR actualiz* [tw] OR "bringing about" [tw] OR complete[tw] OR completes [tw] OR completed[tw] OR completion[tw] OR finish[tw] OR finishes[tw] OR finished[tw] OR perform[tw] OR performs[tw] OR performed[tw] OR performance[tw] OR realize[tw] OR realizes[tw] OR realized[tw] OR realization[tw] OR
Reflecting and evaluating	reflecting [tw] OR evaluating [tw] OR evaluation [tw] OR "quantitative feedback" OR "qualitative feedback" OR "quality of implementation" OR "summative evaluation" OR reflecting OR evaluating
Embase electronic database	
Implementation	(adopt: ab, ti OR adopts: ab, ti OR adopted: ab, ti OR adopting: ab, ti OR adoption: ab, ti NOT 'adoption'/exp) OR implement: ab, ti OR implements: ab, ti OR implementation: ab, ti OR 'health care planning'/exp OR 'quality improvement': ab, ti OR 'total quality management'/exp OR diffused: ab, ti OR diffusion: ab, ti OR 'mass communication'/exp OR 'medical information system'/exp OR 'knowledge translation*': ab, ti OR 'knowledge exchange*': ab, ti AND
Measurement	instrument*': ab, ti OR survey*': ab, ti OR questionnaire*': ab, ti OR 'questionnaire'/exp OR measurement*': ab, ti OR measure*': ab, ti OR inventory: ab, ti AND
Evidence-based practice	'empirically supported treatment' OR 'evidence based practice' OR 'evidence based treatment' OR 'evidence based practice'/exp OR innovation OR 'practice guideline'/exp OR guideline: ab, ti AND

(Continued)



**Table 2.** (Continued)

Pubmed electronic database

Search term	Search string
Behavioral health	'mental health': ab, ti OR 'behavioral health': ab, ti OR psychiatr*: ab, ti OR 'mental disease'/exp OR 'psychiatry'/exp OR 'behavioral medicine'/exp OR 'mental health service'/exp OR 'mental hospital'/exp OR 'psychiatric nursing'/exp AND
Structural characteristics	'structural characteristic*' OR 'social architecture' OR 'age of organization' OR 'maturity of organization' OR 'size of organization' OR 'functional differentiation' OR 'administrative intensity' OR centralization OR
Networks and communication	networks OR communication OR communicate OR communicates OR 'social networks' OR 'formal communication' OR 'informal communication' OR 'internal bonding' OR teamness OR 'shared vision' OR 'information sharing' OR
Planning	planning OR 'health care planning'/exp OR 'comprehensive approach*' OR 'implementation plan*' OR 'tailored implementation*' OR strategies OR strategy OR workaround* OR formulate OR formulation OR formulated OR 'course*' of action' OR 'implementation plans' OR 'tailored implementation' OR strategies OR workarounds OR formulate OR
Engaging	engaged OR engages OR engagement OR attract* OR involv* OR 'social market*' OR model OR models OR 'theoretical model'/exp OR
Opinion leaders	'opinion leaders' OR
Formally appointed internal implementation leaders	'implementation leader*' OR coordinator OR 'project manager*' OR 'team leader*' OR
Champions	champion OR champions OR championed OR 'transformational leader*' OR campaigner* OR promoter* OR proponent* OR supporter* OR
External change agents	'external change agent' OR consultant* OR 'technical assist*' OR facilitator OR
Executing	execute OR executed OR execution OR 'carr*' out' OR accomplish* OR actualiz* OR 'bringing about' OR complete OR completes OR completed OR completion OR finish OR finishes OR finished OR perform OR performs OR performed OR performance OR realize OR realizes OR realized OR realization OR
Reflecting and evaluating	reflecting OR evaluating OR evaluation OR 'quantitative feedback' OR 'qualitative feedback' OR 'quality of implementation' OR 'summative evaluation' OR reflecting OR evaluating

full 122-item Community Leader Survey (Valente et al., 2007), created by adapting multiple existing measures of coalition functioning, planning, and adoption for use in evidence-based drug prevention programs.

*Engaging* ( $n=1$  measure). The *Engaging* construct is further delineated to specify four different actors critical to attract and involve in implementation and two (i.e., *Opinion leaders*, *Champions*) revealed relevant measures. Psychometric data on the single *Engaging* measure, along with measures for the associated *Opinion leaders* and *Champions*, can be found in Table 10. The *Engaging* measure was used in a single study testing the association between patient activation and remission rates (Whitebird et al., 2014) and was created specifically for the medical group surveyed. Only predictive

validity information was available, receiving a “2—adequate” rating.

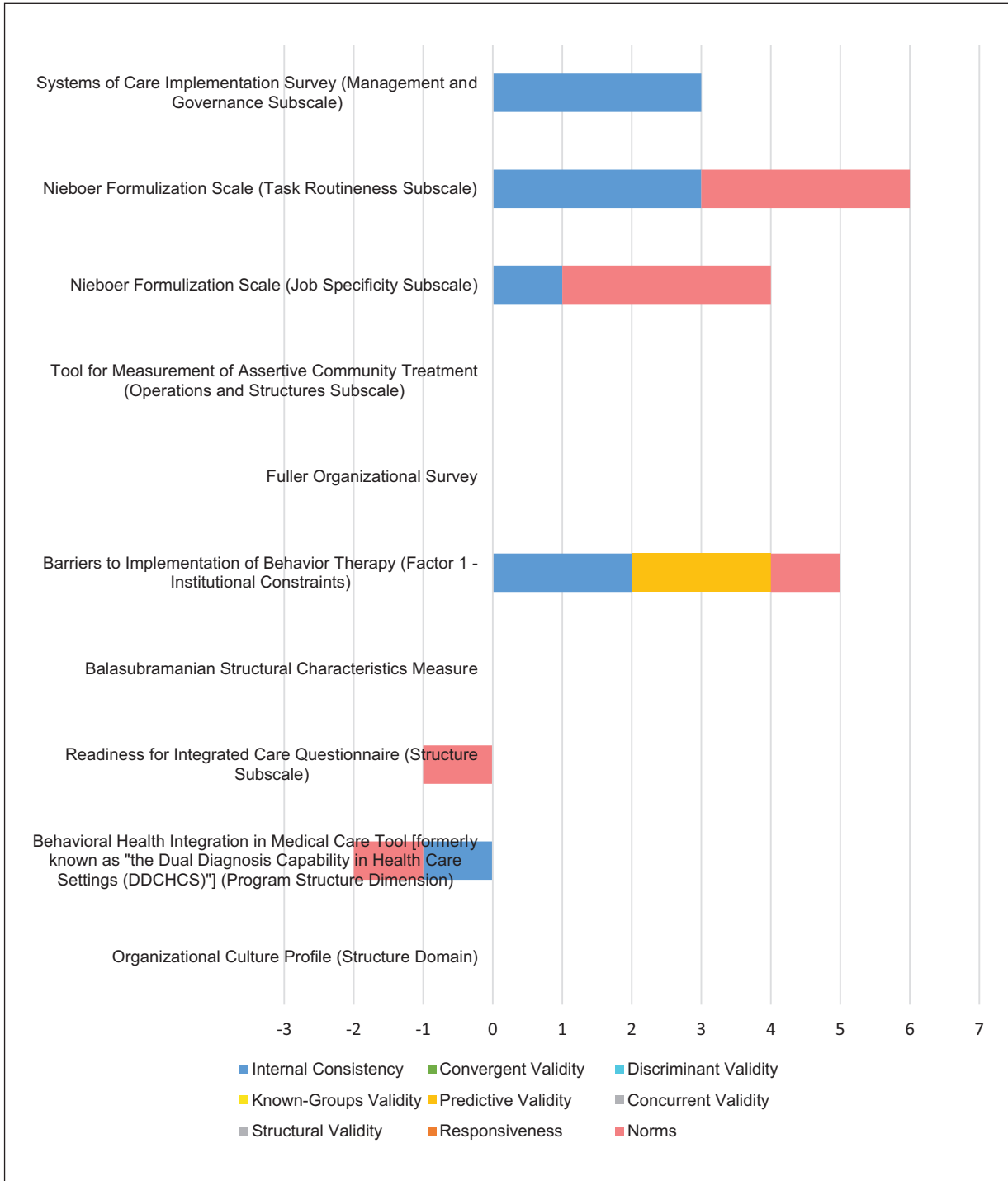
*Opinion leaders* ( $n=5$  measures and scales). More than half of *Opinion leader* measures (Farley et al., 2014; Moore et al., 2004; Ravitz et al., 2013) ( $n=3$ ; 60%) could not be rated using the PAPERS. For instance, to identify opinion leaders, health care professionals filled out a form asking “Which members of your immediate team, with whom you work day-to-day, have you sought advice from, or given advice to on the management of schizophrenia?,” along with the job roles, direction of contact (i.e., gives or receives advice), frequency, and communication mode for these individuals (Farley et al., 2014). Psychometric information was only available for an individual scale from the full 77-item Organizational Readiness for Change

**Table 3.** Characteristics of measures and scales.

	Structural characteristics (N = 10)		Networks and communications (N = 24)		General process (N = 4)		Engaging (N = 1)		Opinion leaders (N = 2)		Champions (N = 3)		Planning (N = 5)		Reflecting and evaluating (N = 6)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Concept defined																
Yes	10	100%	22	92%	4	100%	1	100%	2	100%	3	100%	5	100%	6	100%
No	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
One-time use only																
Yes	8	80%	11	46%	1	25%	1	100%	0	0%	2	67%	2	40%	3	50%
No	2	20%	13	54%	3	75%	0	0%	2	100%	1	33%	3	60%	3	50%
Number of items																
1–5	5	50%	13	54%	2	50%	0	0%	2	100%	2	67%	2	40%	1	17%
6–10	0	0%	5	21%	0	0%	1	100%	0	0%	0	0%	2	40%	1	17%
11 or more	1	10%	3	13%	2	50%	0	0%	0	0%	0	0%	1	20%	4	67%
Not reported	4	40%	3	13%	0	0%	0	0%	0	0%	1	33%	0	0%	0	0%
Country of origin																
United States	7	70%	17	71%	4	100%	1	100%	2	100%	3	100%	4	80%	5	83%
Other	3	30%	7	29%	0	0%	0	0%	0	0%	0	0%	1	20%	1	17%
Setting administered																
State mental health	1	10%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Inpatient psychiatry	2	20%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Outpatient community	6	60%	15	63%	1	25%	0	0%	1	50%	3	100%	3	60%	3	50%
School mental health	1	10%	1	4%	1	25%	0	0%	0	0%	0	0%	1	20%	0	0%
Residential care	2	20%	5	21%	0	0%	0	0%	1	50%	0	0%	0	0%	0	0%
Other	6	60%	16	67%	2	50%	1	100%	1	50%	2	67%	4	80%	3	50%
Level of analysis																
Consumer	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Organization	1	10%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Clinic/site	0	0%	3	13%	1	25%	1	100%	1	50%	2	67%	1	20%	2	33%
Provider	8	80%	20	83%	3	75%	0	0%	2	100%	2	67%	4	80%	3	50%
System	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Team	1	10%	1	4%	0	0%	0	0%	1	50%	1	33%	1	20%	1	17%
Director	1	10%	12	50%	1	25%	0	0%	2	100%	3	100%	1	20%	1	17%
Supervisor	1	10%	12	50%	1	25%	0	0%	1	50%	2	67%	2	40%	1	17%
Other	1	10%	0	0%	1	25%	0	0%	0	0%	0	0%	2	40%	2	33%
Population assessed																
General mental health	2	20%	8	33%	1	25%	0	0%	1	50%	2	67%	2	40%	4	67%
Anxiety	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Depression	0	0%	3	13%	0	0%	1	100%	0	0%	1	33%	0	0%	1	17%
Suicidal ideation	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	17%
Alcohol use disorder	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Substance use disorder	2	20%	11	46%	2	50%	0	0%	1	50%	1	33%	2	40%	3	50%
Behavioral disorder	0	0%	1	4%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Mania	0	0%	2	8%	0	0%	0	0%	0	0%	0	0%	1	20%	0	0%
Eating disorder	0	0%	1	4%	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	100%

(Continued)





**Figure 1.** Structural characteristics: head-to-head comparison of measures or scales.

Assessment (ORCA; Helfrich et al., 2009) and used a five-point scale to assess participant’s level of agreement with four different behaviors exhibited by opinion leaders. Internal consistency received an “4—excellent” score, whereas known-groups validity and norms rated “-1—poor” (Table 10).

*Champions* ( $n=5$  measures and scales). Two measures could not be evaluated using the PAPERS (Dorsey et al., 2016; Guerrero et al., 2015). For example, one study created a dichotomous variable to code the presence or absence of an established local clinical champion of depression screening and treatment in primary care



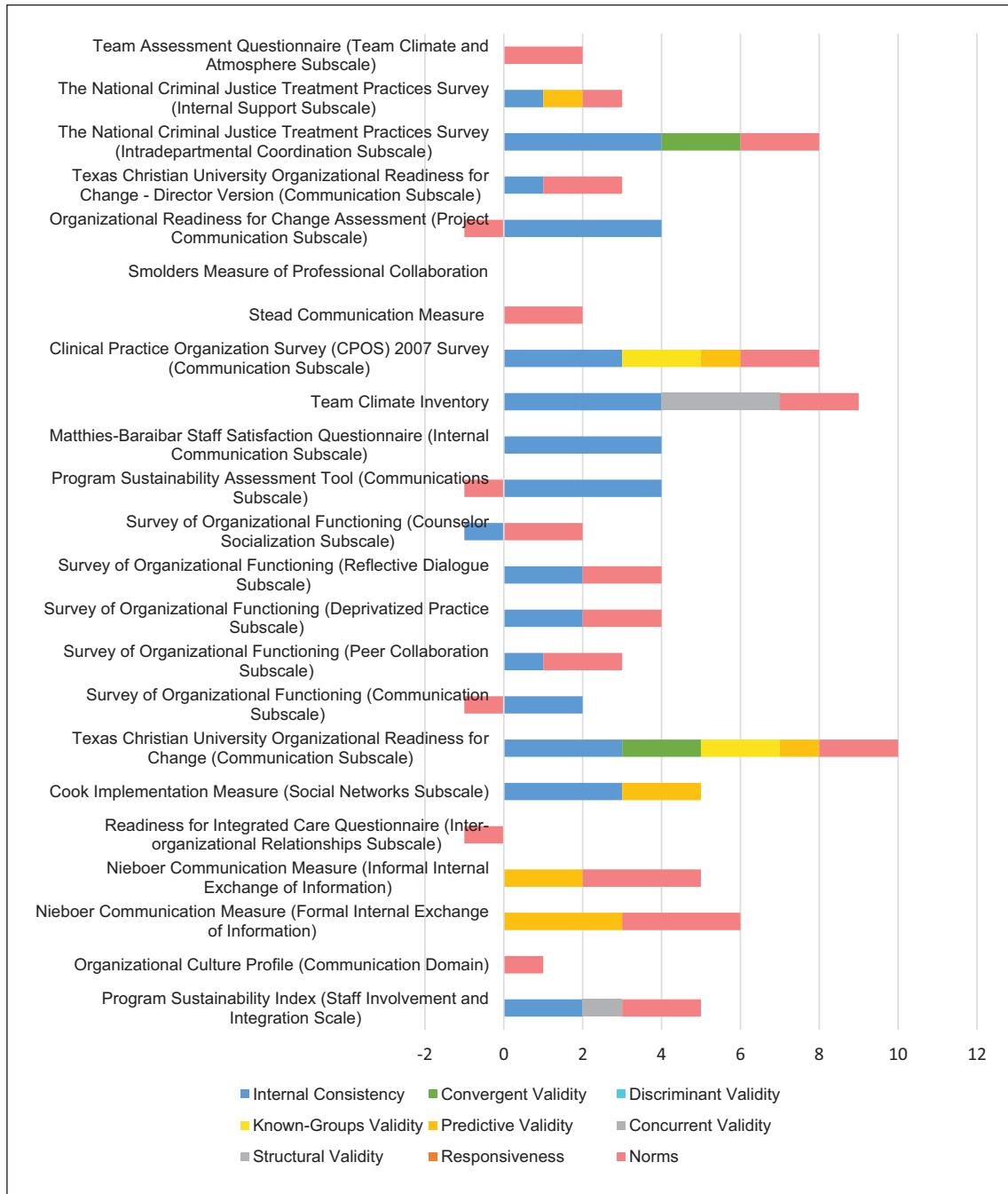
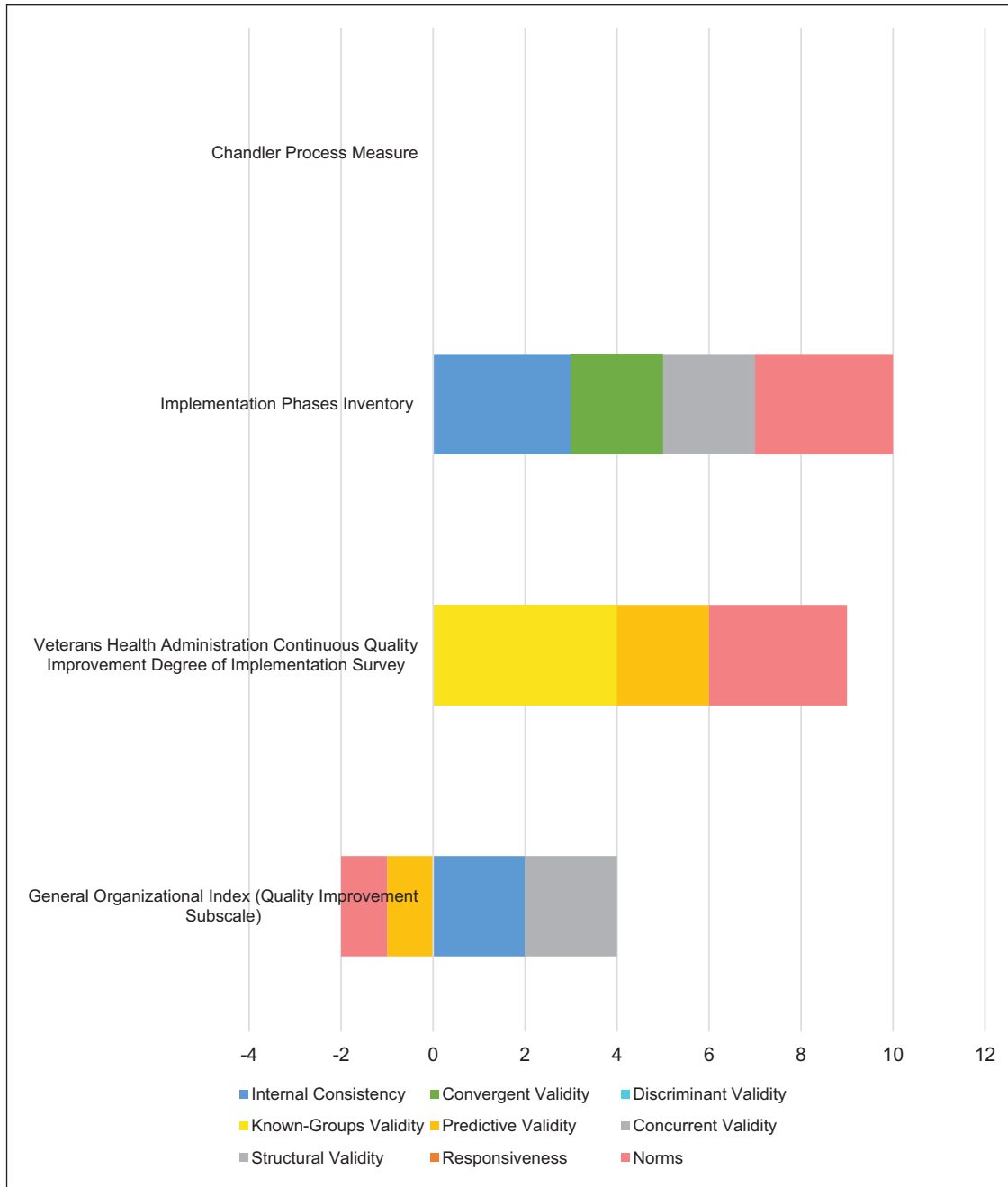


Figure 2. Networks and communications: head-to-head comparison of measures or scales.

(Guerrero et al., 2015). All three rated scales were part of multi-scaled measures and contained psychometric information. Total PAPERS scores ranged from -1 to 5 (Table 10). Information was available for internal consistency ( $n=3$ ; 100%), known-groups validity ( $n=1$ ; 33.33%), predictive validity ( $n=1$ ; 33.33%), and norms ( $n=3$ ; 100%).

Median scores for internal consistency were “4—excellent” and “1—minimal/emerging” for known-groups validity, predictive validity, and norms.

*Reflecting and evaluating* ( $n=5$  measures and scales). Only quantitative assessments of *Reflecting and evaluating* were included. Most reported PAPERS-relevant



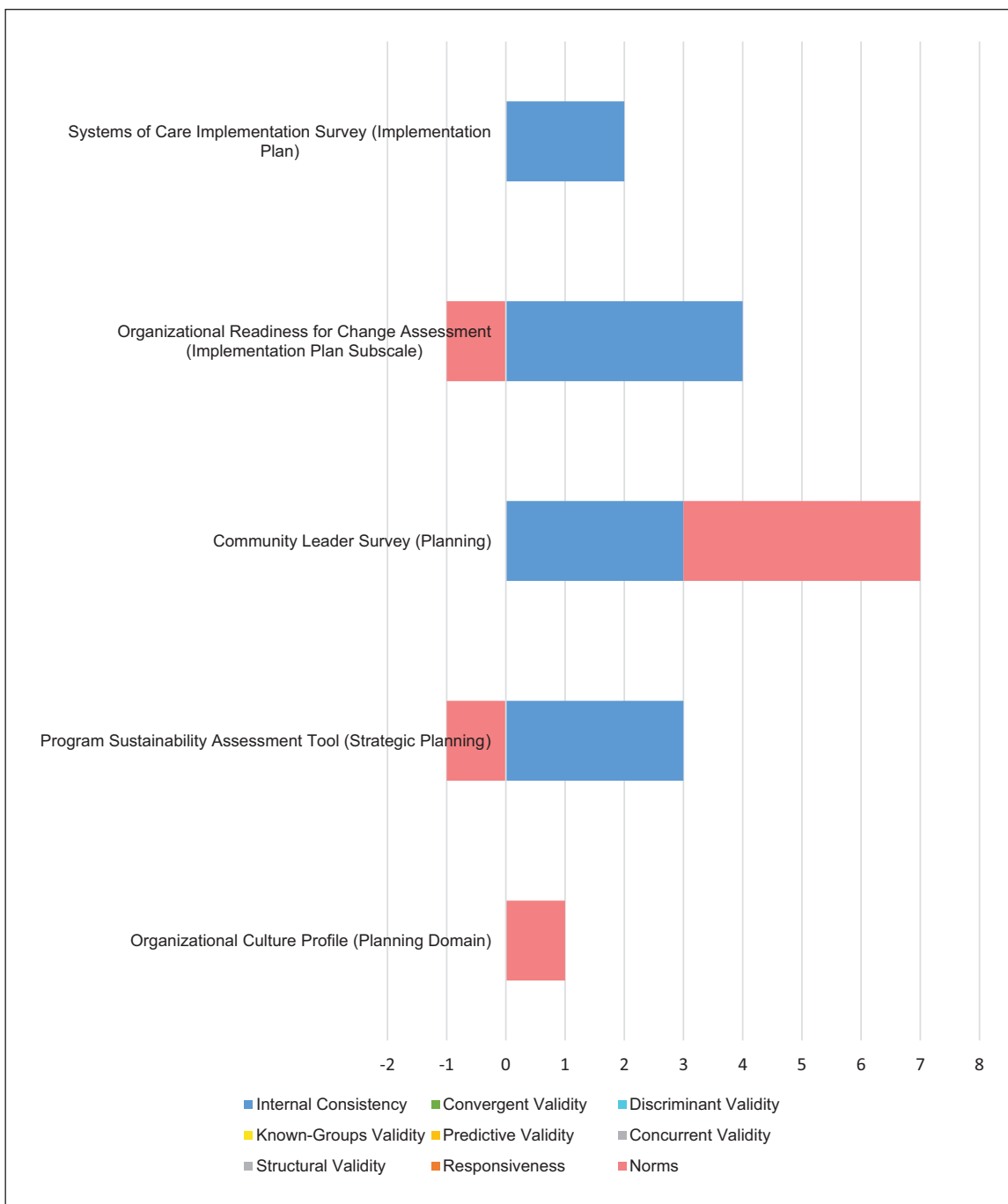
**Figure 3.** General process: head-to-head comparison of measures or scales.

information ( $n=3$ ; 60%), with total PAPERS scores ranging from 2 to 8. Internal consistency ( $n=3$ ; 60%) reported a “3—good” median rating and norms ( $n=3$ ; 60%) contained a median rating of “-1—poor.” In one instance where no psychometric data was available, a study-specific measure using quantitative and qualitative response formats evaluated government and community agencies ( $n=18$ ) progress implementing a 55-activity suicide

prevention strategy (Sheehan et al., 2015). Table 11 describes psychometric properties for each *Reflecting and evaluating* measure.

**Discussion**

Results from this systematic review of constructs central to the implementation process largely replicate those that



**Figure 4.** Planning: head-to-head comparison of measures or scales.

have come before: reporting on key psychometric properties necessary for evaluating the quality of our measurements remains limited. To aide in improving measurement, the PAPERS rating system was developed. Psychometric quality was evaluated using nine distinct properties rated on a scale of “poor” (-1), “none” (0), “minimal/emerging” (1), “adequate” (2), “good” (3), or “excellent” (4). Measures had a potential total score range of -9 to 36, with

higher scores indicating higher psychometric quality. Measurement reviews have been criticized for lacking comprehensive details on psychometric properties; to address this criticism we took an expansive approach by evaluating nine key properties. When comparing to another CFIR-guided measure review evaluating six properties (Clinton-McHarg et al., 2016), only three (i.e., structural validity, internal consistency, responsiveness) overlapped

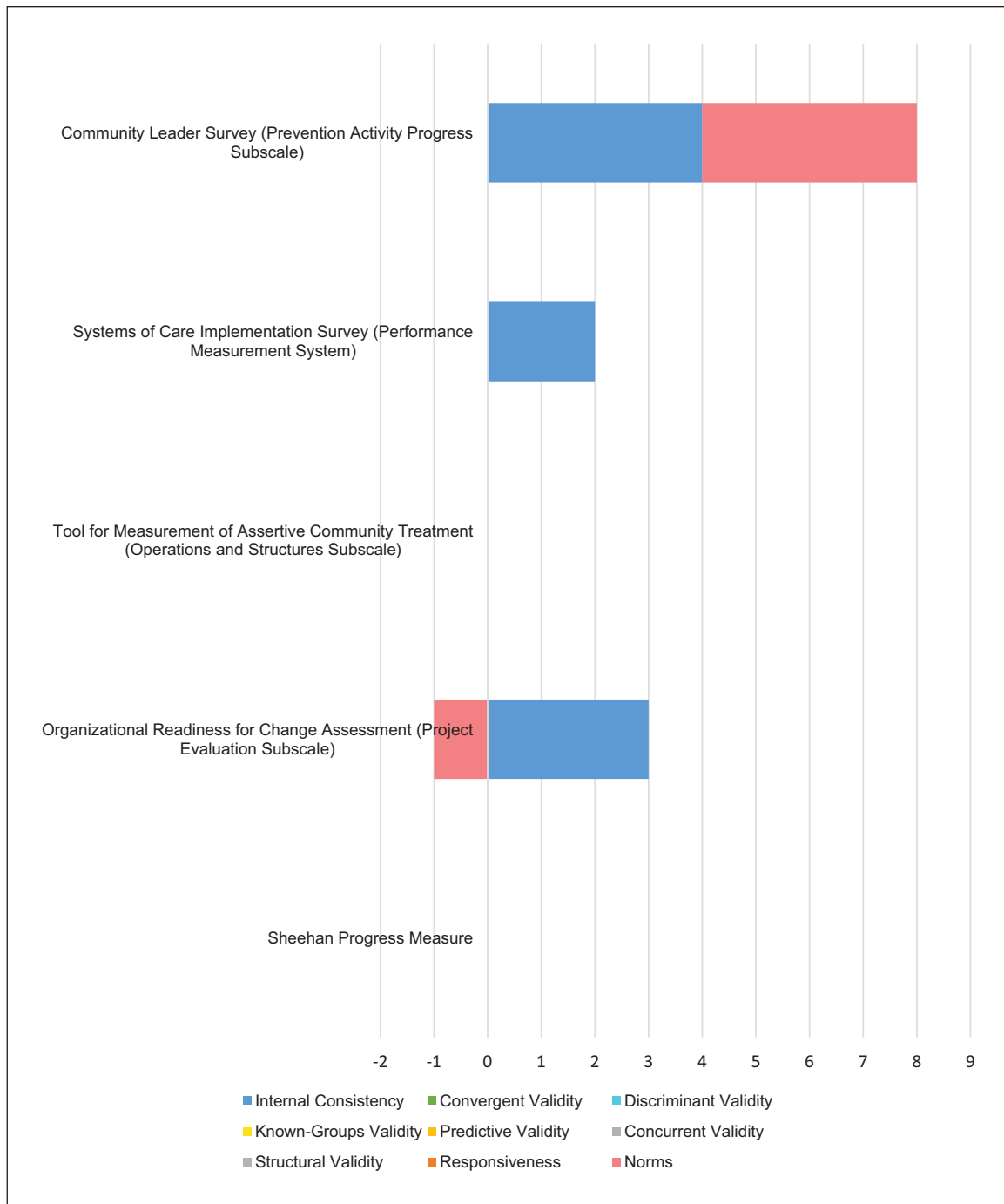


**Figure 5.** Engaging, opinion leaders, and champions: head-to-head comparison of measures or scales.

with the PAPERS. In addition, we attempted to locate all associated literature for each measure, whereas other studies analyzed only a single article. We found that the majority of constructs contained study-specific measures (i.e., “one-time use”), which limits their generalizability. What is more, no measures were identified for 3 of the 10 central constructs: *Formally appointed internal implementation*

*leaders*, *External change agents*, and *Executing*. For identified measures, PAPERS ratings were generally quite low, rating “1—minimal/emergent” to “2—adequate.” The highest overall PAPERS score was 10 out of a possible 36 (higher scores indicate higher quality), observed in just two measures. At the property level, only two of the nine reported a “4—excellent” median score: known-groups





**Figure 6.** Reflecting and evaluating: head-to-head comparison of measures or scales.

validity and internal consistency. Interestingly, this was determined by a single rating within each construct (*General process, Opinion leaders, Champions*) across just two different measures (Helfrich et al., 2009; Parker et al., 1999). These results clearly illuminate significant gaps in measuring implementation processes, while simultaneously identify areas for improvement in future measurement work.

### *State of science*

At least three potential reasons could explain the low quality of implementation measurement (Chambers et al., 2018): 1) implementation science, as a discipline, is so new there simply has not been enough time for sufficient measure development, 2) researchers fail to conduct or report psychometric testing, and/or 3) existing measures

**Table 5.** Summary statistics for measure and scale ratings.

Construct	N	Internal consistency		Convergent validity		Discriminant validity		Known-groups validity		Predictive validity		Concurrent validity		Structural validity		Responsiveness		Norms	
		M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R
Networks and communications	24	3	-1,4	2	2,2	-	-	2	2,2	1	1,3	1	1,1	3 <sup>a</sup>	3	-	-	2	-1,3
Structural characteristics	10	2	-1,3	-	-	-	-	-	-	2 <sup>a</sup>	2	-	-	-	-	-	-	1	-1,3
General process	4	2	2,3	2 <sup>a</sup>	2	-	-	4 <sup>a</sup>	4	-1	-1,2	2	2,2	-	-	-	-	3	-1,3
Engaging	1	-	-	-	-	-	-	-	-	2 <sup>a</sup>	2	-	-	-	-	-	-	-	-
Opinion leaders	2	4 <sup>a</sup>	4	-	-	-	-	-1 <sup>a</sup>	-1	-	-	-	-	-	-	-	-	-1 <sup>a</sup>	1
Champions	3	4 <sup>b</sup>	4	-	-	-	-	1 <sup>a</sup>	1	1 <sup>a</sup>	1	-	-	-	-	-	-	1	-1,2
Planning	5	3	2,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1	-1,4
Reflecting and evaluating	5	3	2,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-1,4

<sup>a</sup>Denotes only one rating available.

**Table 6.** Structural characteristics psychometric ratings by measure or scale.

Measure name	Internal consistency	Convergent validity		Discriminant validity	Known-groups validity	Predictive validity		Concurrent validity		Structural validity		Responsiveness		Total score
		M	R			M	R	M	R	M	R	M	R	
Organizational Culture Profile (Structure Domain) (Dark et al., 2015)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Behavioral Health Integration in Medical Care Tool [formerly known as "the Dual Diagnosis Capability in Health Care Settings (DDCHCS)"] (Program Structure Dimension) (McGovern et al., 2012)	-1	0	0	0	0	0	0	0	0	0	0	0	-1	-2
Readiness for Integrated Care Questionnaire (Structure Subscale) (V. C. Scott et al., 2017)	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1
Balasubramanian Structural Characteristics Measure (Balasubramanian et al., 2015)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barriers to Implementation of Behavior Therapy (Factor 1—Institutional Constraints) (Corrigan et al., 1992)	2	0	0	0	0	2	0	0	0	0	0	0	1	5
Fuller Organizational Survey (Fuller et al., 2007)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tool for Measurement of Assertive Community Treatment (Operations and Structures Subscale) (Monroe-DeVita et al., 2011)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nieboer Formualization Scale (Job Specificity Subscale) (Nieboer & Strating, 2012)	1	0	0	0	0	0	0	0	0	0	0	0	3	4
Nieboer Formualization Scale (Task Routineness Subscale) (Nieboer & Strating, 2012)	3	0	0	0	0	0	0	0	0	0	0	0	3	6
Systems of Care Implementation Survey (Management and Governance Subscale) (Boothroyd et al., 2011)	3	0	0	0	0	0	0	0	0	0	0	0	0	3

DDCHCS: Dual Diagnosis Capability in Health Care Settings.



**Table 9.** Planning psychometric ratings by measure or scale.

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Organizational Culture Profile (Planning Domain) (Dark et al., 2015)	0	0	0	0	0	0	0	0	1	1
Program Sustainability Assessment Tool (Strategic Planning) (Luke et al., 2014)	3	0	0	0	0	0	0	0	-1	2
Community Leader Survey (Planning) (Valente et al., 2007)	3	0	0	0	0	0	0	0	4	7
Organizational Readiness for Change Assessment (Implementation Plan Subscale) (Helfrich et al., 2009)	4	0	0	0	0	0	0	0	-1	3
Systems of Care Implementation Survey (Implementation Plan) (Boothroyd et al., 2011)	2	0	0	0	0	0	0	0	0	2

**Table 10.** Engaging, opinion leaders, and champions psychometric ratings by measure or scale.

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Engaging										
Whitebird Process Measure (Whitebird et al., 2014)	0	0	0	0	2	0	0	0	0	2
Opinion Leaders										
Organizational Readiness for Change Assessment (Opinion Leaders Subscale) (Helfrich et al., 2009)	4	0	0	-1	0	0	0	0	-1	2
Cook Implementation Measure (Opinion Leaders Subscale) (J. M. Cook et al., 2012)	0	0	0	0	0	0	0	0	0	0
Champions										
Readiness for Integrated Care Questionnaire (Program Champion Subscale) (V. C. Scott et al., 2017)	0	0	0	0	0	0	0	0	-1	-1
Clinical Practice Organization Survey (CPOS) 2007 Survey (Clinical Champion for Depression Treatment) (Chang et al., 2013)	0	0	0	1	1	0	0	0	2	4
Organizational Readiness for Change Assessment (Clinical Champion Subscale) (Helfrich et al., 2009)	4	0	0	0	0	0	0	0	1	5

CPOS: Clinical Practice Organization Survey.



**Table 11.** Reflecting and evaluating psychometric ratings by measure or scale.

Measure name	Internal consistency	Convergent validity	Discriminant validity	Known-groups validity	Predictive validity	Concurrent validity	Structural validity	Responsiveness	Norms	Total score
Sheehan Progress Measure (Sheehan et al., 2015)	0	0	0	0	0	0	0	0	0	0
Organizational Readiness for Change Assessment (Project Evaluation Subscale) (Helfrich et al., 2009)	3	0	0	0	0	0	0	0	-1	2
Tool for Measurement of Assertive Community Treatment (Operations and Structures Subscale) (Monroe-DeVita et al., 2011)	0	0	0	0	0	0	0	0	0	0
Systems of Care Implementation Survey (Performance Measurement System) (Boothroyd et al., 2011)	2	0	0	0	0	0	0	0	0	2
Community Leader Survey (Prevention Activity Progress Subscale) (Valente et al., 2007)	4	0	0	0	0	0	0	0	4	8

perform poorly. We observed all three. While just over half of identified measures were developed within the past decade, no single measure contained psychometric evidence for all PAPERS properties. Across two constructs (*General process* and *Opinion leaders*), predictive validity, known-groups validity, and norms reported median ratings of “-1—poor.” What’s more, when analyzing the reported ranges for the norms data, a rating of “-1—poor” was the lowest value across all constructs, except *Engaging*, where no norms data was reported. The PAPERS require reporting on measures of central tendency (i.e., mean and standard deviation) and sample size for rating, a very basic step when conducting a study and reporting results. Sample size is the main component of this PAPERS criterion, the lower the sample size the lower the rating. Both the lack of reporting and small sample sizes we observed are particularly concerning given that the number of participants plays a key role in defining the validity and applicability of a study’s results. We encourage future studies, at the very minimum, to prioritize including norms-related data when reporting results.

The field of implementation science is in need of more precise definitions for the constructs we intend to measure. While we chose the CFIR as our guiding framework due to its comprehensiveness, it can still benefit from more conceptual delineation, especially the three constructs where no measures were identified: *Formally appointed implementation leader*, *External change agents*, and *Executing*. In another study using the CFIR, 102 staff members (e.g., prevention program coordinators, clinical managers and administrators, primary care providers, nurse care managers) across 12 Veterans Health Administration facilities were interviewed to explore what constructs were associated with successful implementation of the “Telephone Lifestyle Coaching” program (Damschroder et al., 2017). Unfortunately, no data emerged for the *External change agents* or *Executing* constructs. However, interviews revealed that having both enthusiastic and capable program implementation leaders (i.e., *Formally appointed implementation leader*) in place before the effort began had a strong impact on program implementation, where the outcome was increasing referral rates. In fact, three of four highest-referral facilities had at least two program implementation leaders actively driving program implementation activities. While qualitative (e.g., interviews) work reveals information on the *Formally appointed implementation leader* construct, such evidence is still lacking in the quantitative realm, further demonstrating the pressing need to operationalize all key concepts for which no measures were identified. From this same study, no data emerged for the *Opinion Leaders* construct. In terms of *Champions*, no clinical champions were formally identified and there was little evidence of staff displaying championing behaviors. Digging deeper into the *Champions* construct, a review of 199 health care studies defining the

champion construct revealed massive inconsistencies; in fact, 37 unique terms were identified (Miech et al., 2018). Most significant, the terms “opinion leader” and “champion” were often used in the same sentence, referring to the same role. This was observed in our sample: one scale (Helfrich et al., 2009) asked respondents to rate their agreement with: “Project Clinical Champion is considered a clinical opinion leader.” Furthermore, no study ( $n=199$ ) analyzed in the review developed or validated a standardized measure that could identify champions, determine their effectiveness, or distinguish between the multitude of types (Miech et al., 2018). Our study mirrored this finding. Without fixed, universal definitions, understanding how and why these roles are key factors to the implementation process remains difficult. Finally, while the *Structural Characteristics* construct is broad by design, this generality has clear implications on measure performance. The CFIR includes “social architecture” (i.e., coordinating independent actions when clustering large numbers of people into smaller groups) and qualitative research reveals how changes in group differentiation negatively impact implementation, specifically how merging organizations, addition and/or elimination of roles, and shift changes create isolation from support structures and increase pressures from new responsibilities (Barwick et al., 2020). One *Structural characteristics* measure in our sample assessed the quality of team’s organization, routines, and collaborations with other agencies and found when collaboration and routine development improved, fidelity scores increased across time (Bergmark et al., 2018); unfortunately the measure scored a 0 on the PAPERS, posing limitations on this finding. Taken together, further delineation of all CFIR constructs would decrease confusion and conflation, leading to improvements in quantitative measure performance and greater confidence in our findings. While this kind of conceptual work was beyond the scope of the current study, attention in future work is imperative.

### *Moving beyond the CFIR constructs*

We conducted an item-level analysis to confirm each measure’s construct assignments. We discovered that many measures mapping to *General process* derived from the Quality Improvement (QI) literature. QI methods emphasize communication, engagement, and participation by everyone involved (Pronovost et al., 2005). Specifically, when organizations self-identify as clinical microsystems (i.e., an interdependent group of people with the capacity to make changes), this creates positive, supportive work environments that encourage building knowledge and taking action, leading to strategic and sustainable improvement through measurement and performance feedback (Batalden et al., 2003). One *General process* measure identified as deriving from QI included process and

outcome monitoring items (Bond et al., 2009), relating to clinical microsystems and the notion of evaluating improvement through social networks.

Furthermore, QI collaboratives (QIC) bring together groups of practitioners from different organizations; a typical QIC feature outlines how individuals affiliated with an outside entity stimulate improvement through sharing experiences from their local setting (Øvretveit et al., 2002). During a national evaluation of six types of QICs in the Netherlands, a theory-driven measure was developed and items relating to external change agents were one of three included dimensions (e.g., external change agents made goal and clarified way to achieve it; (Duckers et al., 2008). Basic psychometric testing was conducted: principal component analysis revealed a 4-item “External change agent support” factor and internal consistency reported acceptable reliability ( $\alpha=0.77$ ). This measure, along with its theoretical framework, can further delineate the *External change agents* construct. In QI work forming implementation teams are part of the “process.” Team functioning and composition not currently captured by the CFIR. The Team Assessment Questionnaire (TAQ; Mahoney et al., 2012), identifies team strengths and weaknesses to inform planning and evaluating team performance, but was ultimately excluded from our sample. The TAQ informs how to improve communication and teamwork, factors described as the basis for developing and sustaining effective teams. Taken together, existing QI conceptual and measurement work on team functioning, team composition, and external change agent support outline critical facets of the implementation process that could be integrated and tested in the next iteration of the CFIR framework.

### *Limitations*

Although we employed a rigorous systematic literature review and rating methodology, there are limitations worth noting. First, our behavioral health scope overlooked other settings where psychometric information was reported. This focus may limit applicability outside of behavioral health settings. Second, the CFIR taxonomy guided locating relevant measures and generating search strings. We made our best attempts with our included synonyms to cast as of a broad net as possible. Despite our efforts, no measures were identified for some constructs. Third, although the PAPERS is comprehensive by covering nine distinct psychometric properties, other relevant properties (e.g., inter-rater and test–retest reliability) were excluded due to lacking literature providing clear, statistically significant cut-off points. Fourth, databases have limited capability for identifying quantitative implementation science measures. “Implementation science” was only recently added (January 2019) as a Medical Subject Headings (MeSH) term; MeSH terms index articles and are key components of search strings. To attempt to boost our yield, we created

10 construct-specific search strings. Fifth, the statistical analysis only included measures that could be evaluated psychometrically.

### Future directions

One question the field of implementation science can begin to grapple with is whether these constructs are even amenable to measures that would benefit from psychometric testing. Our review identified a total of 14 measures deemed “not suitable for rating”: *Networks and communications* ( $n=5$ ), *Structural characteristics* ( $n=4$ ), *Opinion leaders* ( $n=3$ ), and *Champions* ( $n=2$ ). Although these measures were not included in our statistical analysis, we know they are necessary for identifying champions, opinion leaders, and other change agents. Future work investigating methods for how to assess the quality of these atypical measures is needed to ensure consistent measurement of what they intend to measure.

Congruent with our findings, measures are continually developed to fit study-specific needs and do not undergo psychometric testing. The item language found in context-, population-, and/or intervention-specific measures limits their generalizability, creating challenges when trying to synthesize results across studies. Broader item language has the potential for more rigorous psychometric testing and widespread use. One such measure, the Evidence-based Practice Attitudes Scale (EBPAS; (Aarons, 2004), has established national norms (Aarons et al., 2010) and strong psychometric evidence across a variety of settings and samples (C. R. Cook et al., 2018; Rye et al., 2017). It is our hope to accelerate the use of measures like the EBPAS (i.e., not context, population, or intervention-specific), as this will place the field in a better position of developing generalizable knowledge.

To decrease the development of these study-specific measures and supplement existing systematic reviews, the field of implementation science is in need of decision-making tools (Martinez et al., 2014). Given implementation science’s interdisciplinary nature, the surge in measure development can potentially overwhelm researchers and practitioners alike. The National Cancer Institute (NCI)-funded Grid-Enabled Measures project (GEM) (NCI, 2020), is an example of a decision-making tool that can navigate the vast array of measures currently available. Similarly, our systematic reviews results are available in an interactive format on the SIRC website for society members. We, as a field, need to focus on improving the quality of existing measures that embrace generalizable language.

### Conclusion

Careful measure development is crucial for empirically testing and establishing these constructs’ mechanisms of change and their interactions. The results of this

systematic review demonstrate the need for more rigorous measure development, especially for the three *Process* domain constructs (*Formally appointed internal implementation leaders*, *External change agents*, *Executing*), for which no measures were located. Measures identified in this review had limited psychometric data available, performing “1—minimal/emergent” to “2—adequate.” With the plethora measures currently available, targeted efforts aimed at their refinement should occur, instead of creating new measures that are study-specific, likely to be “shelved” or abandoned due to failing to demonstrate psychometric strength (Lewis & Dorsey, 2020). We strongly encourage authors to identify existing measures for use in their studies. Implementation science, particularly the *Process* constructs, can also learn from both QI theory and measure development.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for 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 both an author of this manuscript and editor and associate editor of the journal, *Implementation Research and Practice*, respectively. Due to this conflict, Dr. Lewis was not involved in the editorial or review process for this manuscript.

### ORCID iDs

Kayne D Mettert  <https://orcid.org/0000-0003-1750-7863>  
 Cara C Lewis  <https://orcid.org/0000-0001-8920-8075>

### Supplemental material

Supplemental material for this article is available online.

### References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61–74. <https://doi.org/10.1023/b:mhsr.0000024351.12294.65>
- Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. National norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychological Assessment, 22*(2), 356–365. <https://doi.org/10.1037/a0019188>
- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4–23. <https://doi.org/10.1007/s10488-010-0327-7>

- American Psychological Association. (2020). *Construct*. <https://dictionary.apa.org/construct>
- 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. [https://doi.org/10.1002/\(SICI\)1099-1379\(199805\)19:3<235::AID-JOB837>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1099-1379(199805)19:3<235::AID-JOB837>3.0.CO;2-C)
- Balasubramanian, B. A., Cohen, D. J., Davis, M. M., Gunn, R., Dickinson, L. M., Miller, W. L., Crabtree, B. F., & Stange, K. C. (2015). Learning Evaluation: Blending quality improvement and implementation research methods to study healthcare innovations. *Implementation Science, 10*, 31. <https://doi.org/10.1186/s13012-015-0219-z>
- Barwick, M., Barac, R., Kimber, M., Akrong, L., Johnson, S. N., Cunningham, C. E., Bennett, K., Ashbourne, G., & Godden, T. (2020). Advancing implementation frameworks with a mixed methods case study in child behavioral health. *Translational Behavioral Medicine, 10*(3), 685–704. <https://doi.org/10.1093/tbm/ibz005>
- Batalden, P. B., Nelson, E. C., Edwards, W. H., Godfrey, M. M., & Mohr, J. J. (2003). Microsystems in health care: Part 9. Developing small clinical units to attain peak performance. *The Joint Commission Journal on Quality and Safety, 29*(11), 575–585. [https://doi.org/10.1016/s1549-3741\(03\)29068-7](https://doi.org/10.1016/s1549-3741(03)29068-7)
- Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., Hurford, M. O., Hadley, T., Barg, F. K., & Walsh, L. M. (2015). Predictors of community therapists' use of therapy techniques in a large public mental health system. *JAMA Pediatrics, 169*(4), 374–382. <https://doi.org/10.1001/jamapediatrics.2014.3736>
- Bergmark, M., Bejerholm, U., Svensson, B., & Markstrom, U. (2018). Complex interventions and interorganisational relationships: Examining core implementation components of assertive community treatment. *International Journal of Integrated Care, 18*(4), 11. <https://doi.org/10.5334/ijic.3547>
- Bond, G. R., Drake, R. E., Rapp, C. A., McHugo, G. J., & Xie, H. (2009). Individualization and quality improvement: Two new scales to complement measurement of program fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(5), 349–357. <https://doi.org/10.1007/s10488-009-0226-y>
- Boothroyd, R. A., Greenbaum, P. E., Wang, W., Kutash, K., & Friedman, R. M. (2011). Development of a measure to assess the implementation of children's systems of care: The Systems of Care Implementation Survey (SOCIS). *The Journal of Behavioral Health Services & Research, 38*(3), 288–302. <https://doi.org/10.1007/s11414-011-9239-x>
- Bradshaw, C. P., Debnam, K., Koth, C. W., & Leaf, P. (2009). Preliminary validation of the implementation phases inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions, 11*(3), 145–160. <https://doi.org/10.1177/1098300708319126>
- 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.jsat.2006.12.030>
- Chambers, D. A., Vinson, C. A., & Norton, W. E. (2018). *Advancing the science of implementation across the cancer continuum*. Oxford University Press.
- Chandler, D. (2009). Implementation of integrated dual disorders treatment in eight California programs. *American Journal of Psychiatric Rehabilitation, 12*(4), 330–351. <https://doi.org/10.1080/15487760903248473>
- 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*, 22. <https://doi.org/10.1186/1748-5908-8-22>
- Chor, K. H., Wisdom, J. P., Olin, S. C., Hoagwood, K. E., & Horwitz, S. M. (2015). Measures for predictors of innovation adoption. *Administration and Policy in Mental Health and Mental Health Services Research, 42*, 545–573. <https://doi.org/10.1007/s10488-014-0551-7>
- Clinton-McHarg, T., Yoong, S. L., Tzelepis, F., Regan, T., Fielding, A., Skelton, E., Kingsland, M., Ooi, J. Y., & Wolfenden, L. (2016). Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: A systematic review. *Implementation Science, 11*(1), 148. <https://doi.org/10.1186/s13012-016-0512-5>
- Cook, C. R., Davis, C., Brown, E. C., Locke, J., Ehrhart, M. G., Aarons, G. A., Larson, M., & Lyon, A. R. (2018). Confirmatory factor analysis of the Evidence-Based Practice Attitudes Scale with school-based behavioral health consultants. *Implementation Science, 13*(1), 116. <https://doi.org/10.1186/s13012-018-0804-z>
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine, 119*(2), 166.e7–166.e16. <https://doi.org/10.1016/j.amjmed.2005.10.036>
- 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*, 59. <https://doi.org/10.1186/1748-5908-7-59>
- Corrigan, P. W., Kwartarini, W. Y., & Pramana, W. (1992). Staff perception of barriers to behavior therapy at a psychiatric hospital. *Behavior Modification, 16*(1), 132–144. <https://doi.org/10.1177/01454455920161007>
- Damschroder, L. J., Aron, D., Keith, R., Kirsh, S., Alexander, J., & Lowery, J. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. <https://doi.org/10.1186/1748-5908-4-50>
- Damschroder, L. J., & Lowery, J. C. (2013). Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). *Implementation Science, 8*, 51. <https://doi.org/10.1186/1748-5908-8-51>
- Damschroder, L. J., Reardon, C. M., Sperber, N., Robinson, C. H., Fickel, J. J., & Oddone, E. Z. (2017). Implementation evaluation of the Telephone Lifestyle Coaching (TLC) program: Organizational factors associated with successful



- implementation. *Translational Behavioral Medicine*, 7(2), 233–241. <https://doi.org/10.1007/s13142-016-0424-6>
- 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(1), Article 310. <https://doi.org/10.1186/s12913-015-0953-6>
- DeVellis, R. F. (2003). *Scale development: Theory and applications*. SAGE.
- Dorsey, S., Berliner, L., Lyon, A. R., Pullmann, M. D., & Murray, L. K. (2016). A statewide common elements initiative for children's mental health. *The Journal of Behavioral Health Services & Research*, 43(2), 246–261. <https://doi.org/10.1007/s11414-014-9430-y>
- Duckers, M. L., Wagner, C., & Groenewegen, P. P. (2008). Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives. *BMC Health Services Research*, 8, Article 172. <https://doi.org/10.1186/1472-6963-8-172>
- Emmons, K. M., Weiner, B., Fernandez, M. E., & Tu, S. P. (2012). Systems antecedents for dissemination and implementation: A review and analysis of measures. *Health Education & Behavior*, 39(1), 87–105. <https://doi.org/10.1177/1090198111409748>
- Farley, K., Hanbury, A., & Thompson, C. (2014). Gathering opinion leader data for a tailored implementation intervention in secondary healthcare: A randomised trial. *BMC Medical Research Methodology*, 14(1), Article 38. <https://doi.org/10.1186/1471-2288-14-38>
- Feinberg, M. E., Riggs, N. R., & Greenberg, M. T. (2005). Social networks and community prevention coalitions. *Journal of Primary Prevention*, 26(4), 279–298. <https://doi.org/10.1007/s10935-005-5390-4>
- Flogdren, G., O'Brien, M. A., Parmelli, E., & Grimshaw, J. M. (2019). Local opinion leaders: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*, 6, CD000125. <https://doi.org/10.1002/14651858.CD000125.pub5>
- Fuller, B. E., Rieckmann, T., Nunes, E. V., Miller, M., Arfken, C., Edmundson, E., & McCarty, D. (2007). Organizational readiness for change and opinions toward treatment innovations. *Journal of Substance Abuse Treatment*, 33(2), 183–192. <https://doi.org/10.1016/j.jsat.2006.12.026>
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly*, 82, 581–629. <https://doi.org/10.1111/j.0887-378X.2004.00325.x>
- Guerrero, E. G., Heslin, K. C., Chang, E., Fenwick, K., & Yano, E. (2015). Organizational correlates of implementation of colocation of mental health and primary care in the Veterans Health Administration. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(4), 420–428. <https://doi.org/10.1007/s10488-014-0582-0>
- Hanbury, A. (2013). The relative influence of team climate, team norms and social network norms on health professionals' implementation of a national recommendation to offer service-users diagnosed with schizophrenia family intervention therapy. *Psychology, Health & Medicine*, 18(5), 619–625. <https://doi.org/10.1080/13548506.2013.764459>
- Hanna, J., Kubiak, S., Pasmán, E., Gaba, A., Andre, M., Smelson, D., & Pinals, D. A. (2020). Evaluating the implementation of a prisoner re-entry initiative for individuals with opioid use and mental health disorders: Application of the consolidated framework for implementation research in a cross-system initiative. *Journal of Substance Abuse Treatment*, 108, 104–114. <https://doi.org/10.1016/j.jsat.2019.06.012>
- 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(1), 38. <https://doi.org/10.1186/1748-5908-4-38>
- Howell, J. M., & Higgins, C. A. (1990). Champions of change: Identifying, understanding, and supporting champions of technological innovations. *Organizational Dynamics*, 19(1), 40–55. [https://doi.org/10.1016/0090-2616\(90\)90047-S](https://doi.org/10.1016/0090-2616(90)90047-S)
- Hrisos, S., Eccles, M. P., Francis, J. J., Dickinson, H. O., Kaner, E. F., Beyer, F., & Johnston, M. (2009). Are there valid proxy measures of clinical behaviour? A systematic review. *Implementation Science*, 4, 37. <https://doi.org/10.1186/1748-5908-4-37>
- Ibrahim, S., & Sidani, S. (2015). Fidelity of intervention implementation: A review of instruments. *Health Affairs*, 7(12), 1687–1695. <https://doi.org/10.4236/health.2015.712183>
- Johnson, K., Quanbeck, A., Maus, A., Gustafson, D. H., & Dearing, J. W. (2015). Influence networks among substance abuse treatment clinics: Implications for the dissemination of innovations. *Translational Behavioral Medicine*, 5(3), 260–268. <https://doi.org/10.1007/s13142-015-0327-y>
- King, T., & Byers, J. F. (2007). A review of organizational culture instruments for nurse executives. *The Journal of Nursing Administration*, 37(1), 21–31. <https://doi.org/10.1097/00005110-200701000-00005>
- Lehman, W. E. K., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, 22(4), 197–209. [https://doi.org/10.1016/S0740-5472\(02\)00233-7](https://doi.org/10.1016/S0740-5472(02)00233-7)
- Lewis, C. C., & Dorsey, C. (2020). Advancing implementation science measurement. In B. Albers, A. Shlonsky, & R. Mildon (Eds.), *Implementation science 3.0* (pp. 227–251). Springer.
- 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, 155. <https://doi.org/10.1186/s13012-015-0342-x>
- Lewis, C. C., Mettert, K. D., Dorsey, C. N., Martinez, R. G., Weiner, B. J., Nolen, E., Stanick, C., Halko, H., & Powell, B. J. (2018). An updated protocol for a systematic review of implementation-related measures. *Systematic Reviews*, 7(66). <https://doi.org/10.1186/s13643-018-0728-3>
- Li, J., Weeks, M. R., Borgatti, S. P., Clair, S., & Dickson-Gomez, J. (2012). A social network approach to demonstrate the diffusion and change process of intervention from peer health advocates to the drug using community. *Substance Use & Misuse*, 47(5), 474–490. <https://doi.org/10.3109/10826084.2012.644097>
- Luke, D. A., Calhoun, A., Robichaux, C. B., Elliott, M. B., & Moreland-Russell, S. (2014). The Program Sustainability Assessment Tool: A new instrument for public health programs. *Preventing Chronic Disease*, 11, 130184. <https://doi.org/10.5888/pcd11.130184>

- Lundgren, L., Krull, I., Zerden Lde, S., & McCarty, D. (2011). Community-based addiction treatment staff attitudes about the usefulness of evidence-based addiction treatment and CBO organizational linkages to research institutions. *Evaluation and Program Planning, 34*(4), 356–365. <https://doi.org/10.1016/j.evalprogplan.2011.02.002>
- 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>
- Mancini, J. A., & Marek, L. I. (2004). Sustaining community-based programs for families: Conceptualization and measurement. *Family Relations, 53*(4), 339–347. <http://www.jstor.org/stable/3700173>
- Martinez, R. G., Lewis, C. C., & Weiner, B. J. (2014). Instrumentation issues in implementation science. *Implementation Science, 9*, 118. <https://doi.org/10.1186/s13012-014-0118-8>
- Matthies-Baraibar, C., Arcelay-Salazar, A., Cantero-González, D., Colina-Alonso, A., García-Urbaneja, M., González-Llinares, R. M., Letona-Aranburu, J., Martínez-Carazo, C., Mateos-del Pino, M., Nuño-Solinis, R., Olaetxea-Urizar, E., de la Rica-Giménez, J. A., Rodríguez-González, M. A., & Dabouza-Acebal, S. (2014). Is organizational progress in the EFQM model related to employee satisfaction? *BMC Health Services Research, 14*(1), Article 468. <https://doi.org/10.1186/1472-6963-14-468>
- McGovern, M. P., Urada, D., Lambert-Harris, C., Sullivan, S. T., & Mazade, N. A. (2012). Development and initial feasibility of an organizational measure of behavioral health integration in medical care settings. *Journal of Substance Abuse Treatment, 43*(4), 402–409. <https://doi.org/10.1016/j.josat.2012.08.013>
- Miech, E. J., Rattray, N. A., Flanagan, M. E., Damschroder, L., Schmid, A. A., & Damush, T. M. (2018). Inside help: An integrative review of champions in healthcare-related implementation. *SAGE Open Medicine, 6*, 205031211877326. <https://doi.org/10.1177/2050312118773261>
- Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2011). The TMACT: A new tool for measuring fidelity to assertive community treatment. *Journal of the American Psychiatric Nurses Association, 17*(1), 17–29. <https://doi.org/10.1177/1078390310394658>
- Moore, K. A., Peters, R. H., Hills, H. A., LeVasseur, J. B., Rich, A. R., Hunt, W. M., Young, M. S., & Valente, T. W. (2004). Characteristics of opinion leaders in substance abuse treatment agencies. *The American Journal of Drug and Alcohol Abuse, 30*(1), 187–203. <https://doi.org/10.1081/ada-120029873>
- National Cancer Institute. (2020). *Grid-Enabled Measures Database (GEM)*. <http://www.gem-beta.org>
- Nieboer, A. P., & Strating, M. M. (2012). Innovative culture in long-term care settings: The influence of organizational characteristics. *Health Care Management Review, 37*(2), 165–174. <https://doi.org/10.1097/HMR.0b013e318222416b>
- Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science, 10*(1), 53. <https://doi.org/10.1186/s13012-015-0242-0>
- Øvretveit, J., Bate, P., Cleary, P., Cretin, S., Gustafson, D., McInnes, K., McLeod, H., Molfenter, T., Plsek, P., Robert, G., Shortell, S., & Wilson, T. (2002). Quality collaboratives: Lessons from research. *Quality and Safety in Health Care, 11*, 345–351. <https://doi.org/10.1136/qhc.11.4.345>
- Palinkas, L. A., Holloway, I. W., Rice, E., Fuentes, D., Wu, Q., & Chamberlain, P. (2011). Social networks and implementation of evidence-based practices in public youth-serving systems: A mixed-methods study. *Implementation Science, 6*(1), 113. <https://doi.org/10.1186/1748-5908-6-113>
- Parker, V. A., Wubbenhorst, W. H., Young, G. J., Desai, K. R., & Charns, M. P. (1999). Implementing quality improvement in hospitals: The role of leadership and culture. *American Journal of Medical Quality, 14*(1), 64–69. <https://doi.org/10.1177/106286069901400109>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Pronovost, P., Weast, B., Rosenstein, B., Sexton, J. B., Holzmüller, C. G., Paine, L., Davis, R., & Rubin, H. R. (2005). Implementing and validating a comprehensive unit-based safety program. *Journal of Patient Safety, 1*(1), 33–40. <https://doi.org/10.1097/01209203-200503000-00008>
- Ramsay, P. P., Shortell, S. M., Casalino, L. P., Rodriguez, H. P., & Rittenhouse, D. R. (2016). A longitudinal study of medical practices' treatment of patients who use tobacco. *American Journal of Preventive Medicine, 50*(3), 328–335. <https://doi.org/10.1016/j.amepre.2015.07.005>
- Ravitz, P., Cooke, R. G., Mitchell, S., Reeves, S., Teshima, J., Lokuge, B., Lawson, A., McNaughton, N., Skinner, W., & Cooper, C. (2013). Continuing education to go: Capacity building in psychotherapies for front-line mental health workers in underserved communities. *The Canadian Journal of Psychiatry, 58*(6), 335–343. <https://doi.org/10.1177/070674371305800605>
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). The Free Press.
- Rycroft-Malone, J., Seers, K., Chandler, J., Hawkes, C. A., Crichton, N., Allen, C., Bullock, I., & Strunin, L. (2013). The role of evidence, context, and facilitation in an implementation trial: Implications for the development of the PARIHS framework. *Implementation Science, 8*(1), 28. <https://doi.org/10.1186/1748-5908-8-28>
- Rye, M., Torres, E. M., Friborg, O., Skre, I., & Aarons, G. A. (2017). The Evidence-based Practice Attitude Scale-36 (EBPAS-36): A brief and pragmatic measure of attitudes to evidence-based practice validated in US and Norwegian samples. *Implementation Science, 12*(1), 44. <https://doi.org/10.1186/s13012-017-0573-0>
- Schoenwald, S. K., Chapman, J. E., Kelleher, K., Hoagwood, K. E., Landsverk, J., Stevens, J., Glisson, C., Rolls-Reutz, J., Health, R. N., & o, Y. M. (2008). A survey of the infrastructure for children's mental health services: Implications for the implementation of empirically supported treatments (ESTs). *Administration and Policy in Mental Health and Mental Health Services Research, 35*(1–2), 84–97. <https://doi.org/10.1007/s10488-007-0147-6>
- 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>
- Sheehan, J., Griffiths, K., Rickwood, D., & Carron-Arthur, B. (2015). Evaluating the implementation of “Managing the Risk of Suicide: A Suicide Prevention Strategy for the ACT 2009–2014”: A whole-of-government/whole-of-community suicide prevention strategy. *Crisis*, 36(1), 4. <https://doi.org/10.1027/0227-5910/a000295>
- Siegel, S. (1964). *Decision and choice: Contributions of Sidney Siegel*. McGraw Hill.
- Smolders, M., Laurant, M., Verhaak, P., Prins, M., van Marwijk, H., Penninx, B., Wensing, M., & Grol, R. (2010). Which physician and practice characteristics are associated with adherence to evidence-based guidelines for depressive and anxiety disorders? *Medical Care*, 48(3), 240–248. <https://doi.org/10.1097/MLR.0b013e3181ca27f6>
- Soo, S., Berta, W., & Baker, G. R. (2009). Role of champions in the implementation of patient safety practice change. *Healthcare Quarterly*, 12, 123–128. <https://doi.org/10.12927/hcq.2009.20979>
- Squires, J., Estabrooks, C., O'Rourke, H., Gustavsson, P., Newburn-Cook, C., & Wallin, L. (2011). A systematic review of the psychometric properties of self-report research utilization measures used in healthcare. *Implementation Science*, 6, 83. <https://doi.org/10.1186/1748-5908-6-83>
- 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(1), 11–20. <https://doi.org/10.1093/tbm/ibz164>
- Stead, K., Kumar, S., Schultz, T. J., Tiver, S., Pirone, C. J., Adams, R. J., & Wareham, C. A. (2009). Teams communicating through STEPPS. *Medical Journal of Australia*, 190(Suppl. 11), S128–S132. <https://www.ncbi.nlm.nih.gov/pubmed/19485861>
- Stichter, J. P., & Conroy, M. A. (2004). A critical analysis of the role of measurement on the validity of emotional and behavioral disorders (EBD) research. *Behavioral Disorders*, 30(1), 7–18. <https://doi.org/10.1177/019874290403000106>
- Taxman, F. S., Young, D. W., & Fletcher, B. W. (2007). The National Criminal Justice Treatment Practices survey: An overview of the special issue. *Journal of Substance Abuse Treatment*, 32(3), 221–223. <https://doi.org/10.1016/j.jat.2006.12.017>
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (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, 651–657. <https://doi.org/10.1007/s11136-011-9960-1>
- Valente, T. W. (2010). *Social networks and health: Models, methods, and applications*. Oxford University Press.
- Valente, T. W., Chou, C. P., & Pentz, M. A. (2007). Community coalitions as a system: Effects of network change on adoption of evidence-based substance abuse prevention. *American Journal of Public Health*, 97(5), 880–886. <https://doi.org/10.2105/ajph.2005.063644>
- Weiner, B. J., Amick, H., & Lee, S. Y. (2008). Conceptualization and measurement of organizational readiness for change: A review of the literature in health services research and other fields. *Medical Care Research and Review*, 65(4), 379–436. <https://doi.org/10.1177/1077558708317802>
- Whitebird, R. R., Solberg, L. I., Jaeckels, N. A., Pietruszewski, P. B., Hadzic, S., Unutzer, J., Ohnsorg, K. A., Rossom, R. C., Beck, A., Joslyn, K. E., & Rubenstein, L. V. (2014). Effective implementation of collaborative care for depression: What is needed? *The American Journal of Managed Care*, 20(9), 699–707. <https://www.ncbi.nlm.nih.gov/pubmed/25365745>
- Wood, K., Giannopoulos, V., Louie, E., Baillie, A., Uribe, G., Lee, K. S., Haber, P. S., & Morley, K. C. (2020). The role of clinical champions in facilitating the use of evidence-based practice in drug and alcohol and mental health settings: A systematic review. *Implementation Research and Practice*. Advance online publication. <https://doi.org/10.1177/2633489520959072>