



# Determining the influence of intervention characteristics on implementation success requires reliable and valid measures: Results from a systematic review

Implementation Research and Practice  
Volume 2: Jan-Dec 2021 1–15  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2633489521994197  
journals.sagepub.com/home/irp

Cara C Lewis<sup>1,2</sup> , Kayne Mettert<sup>1</sup>   
and Aaron R Lyon<sup>2</sup>

## Abstract

**Background:** Despite their inclusion in Rogers' seminal diffusion of innovations theory, few implementation studies empirically evaluate the role of intervention characteristics. Now, with growing evidence on the role of adaptation in implementation, high-quality measures of characteristics such as adaptability, trialability, and complexity are needed. Only two systematic reviews of implementation measures captured those related to the intervention or innovation and their assessment of psychometric properties was limited. This manuscript reports on the results of eight systematic reviews of measures of intervention characteristics with nuanced data regarding a broad range of psychometric properties.

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

**Results:** Searches identified 16 measures or scales: zero for *intervention source*, one for *evidence strength and quality*, nine for *relative advantage*, five for *adaptability*, six for *trialability*, nine for *complexity*, and two for *design quality and packaging*. Information about internal consistency and norms was available for most measures, whereas information about other psychometric properties was most often not available. Ratings for psychometric properties fell in the range of "poor" to "good."

**Conclusion:** The results of this review confirm that few implementation scholars are examining the role of intervention characteristics in behavioral health studies. Significant work is needed to both develop new measures (e.g., for intervention source) and build psychometric evidence for existing measures in this forgotten domain.

## Plain Language Summary

Intervention characteristics have long been perceived as critical factors that directly influence the rate of adopting an innovation. It remains unclear the extent to which intervention characteristics including *relative advantage*, *complexity*, *trialability*, *intervention source*, *design quality and packaging*, *evidence strength and quality*, *adaptability*, and *cost* impact implementation of evidence-based practices in behavioral health settings. To unpack the differential influence of these factors, high quality measures are needed. Systematic reviews can identify measures and synthesize the data regarding their quality to identify gaps in the field and inform measure development and testing efforts. Two previous reviews

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

<sup>2</sup>Department of Psychiatry and Behavioral Sciences and Harborview Medical Center, University of Washington, Seattle, WA, USA

## Corresponding author:

Cara C Lewis, Kaiser Permanente Washington Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA 98101, USA.  
Email: cara.c.lewis@kp.org



identified measures of intervention characteristics, but they did not provide information about the extent of the existing evidence nor did they evaluate the host of evidence available for identified measures. This manuscript summarizes the results of nine systematic reviews (i.e., one for each of the factors listed above) for which 16 unique measures or scales were identified. The nuanced findings will help direct measure development work in this forgotten domain.

### Keywords

Intervention characteristics, implementation, dissemination, evidence-based practice, mental health, measurement, reliability, validity, psychometric

In 2009, the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2009) synthesized the implementation theories, frameworks, and models to offer a meta-framework that could guide measurement of key implementation constructs housed within five domains: intervention characteristics, process, characteristics of individuals, inner setting, and outer setting. Of these, the domain of intervention characteristics has received some of the least empirical attention from implementation scholars. This is ironic, given its prominence in Rogers' seminal diffusion of innovations theory (Turzhitsky et al., 2010) and other theoretical frameworks or teaching tools that explicitly call out the importance of aspects of the innovation—or “the thing” (Curran, 2020)—being implemented. Rogers theorized that the perceived attributes of an innovation or intervention directly affect its rate of adoption, including its *relative advantage, compatibility, complexity, and trialability*. This list is expanded in the CFIR to include the following aspects from the extant literature: *intervention source, design quality and packaging, evidence strength and quality, adaptability, and cost*. Interestingly, the CFIR includes *compatibility* in the inner setting domain for a total of eight constructs in the intervention characteristics domain. Despite the long-standing interest in these constructs theoretically, Lyon and Bruns (2019) have recently

referred to intervention characteristics as the forgotten domain of implementation science.

Intervention characteristics may have a differential influence across the stages or phases of implementation (Moullin et al., 2019). For instance, the cost of an intervention might be prohibitive making it so that organizations fail to adopt it in the exploration phase. Trialability might be most salient in the preparation phase as organizations seek to expose providers to the intervention and give them meaningful experiences with it. Complexity might drive strategy selection and deployment in the implementation phase. Finally, relative advantage might explain why some interventions are maintained and others are not, for example. Furthermore, it might be the case that certain intervention characteristics matter more (e.g., adaptability) in resource-poor contexts where the majority of the population receives their mental health care. These possibilities actually remain empirical questions that could critically inform intervention design, adoption decisions, implementation, and sustainment efforts. However, to pursue these types of research questions, the field needs reliable and valid measures of intervention characteristics.

Of the 15 systemic reviews of implementation measures (J. D. Allen et al., 2017; Miake-Lye et al., 2020; Rabin et al., 2016; Willmeroth et al., 2019), only two explored measures of intervention/innovation characteristics (Chaudoir et al.,

**Table 1.** Definitions of intervention characteristics constructs.

Intervention Source	Perception of key stakeholders about whether the intervention is externally or internally developed.
Evidence Strength & Quality	Stakeholders' perceptions of the quality and validity of evidence supporting the belief that the intervention will have desired outcomes.
Relative Advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.
Adaptability	The degree to which an intervention can be adapted, tailored, refined, or reinvented to meet local need.
Trialability	The ability to test the intervention on a small scale in the organization, and to be able to reverse course (undo implementation) if warranted.
Complexity	Perceived intricacy or difficulty of *the innovation*, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy and number of steps required to implement.
Design Quality & Packaging Cost	Perceived excellence in how the intervention is bundled, presented, and assembled. Costs of the intervention and costs associated with implementing that intervention, including investment, supply, and opportunity costs.

Note: These definitions are pulled directly from the Consolidated Framework for Implementation Research (Damschroder et al., 2009).

2013; Chor et al., 2015). These studies made important contributions to the implementation science literature. Together, the studies uncovered 45 measures. Both studies concluded, across the domains they reviewed, that the state of measurement was only emerging, with only 38% of measures across both studies demonstrating any evidence of reliability and validity (45% of those in Chor et al., 2015; 25% of those in Chaudoir et al., 2013). However, these two previous reviews only reported on whether (or not) measures had *any* evidence of reliability or validity (dichotomous yes/no), which makes it difficult for those deciding between measures given the limited information. It would be more useful for those seeking a measure to know the extent or degree to which the measure is reliable or valid, and to have information about different types of psychometric properties (e.g., concurrent validity versus predictive validity). Furthermore, neither study systematically sought all articles reflecting empirical uses of identified measures. Since reliability and validity are to some extent context-dependent, assessing all available evidence would provide a more complete picture of the measure's psychometric properties. Now, 5+ years later, we are in a position to update these most relevant reviews and offer new information given the differences in our protocols, notably a more nuanced assessment of each measure's psychometric evidence across all available studies. This manuscript reports on eight systematic reviews—one for each construct outlined by the CFIR (see definitions in Table 1)—of measures of intervention characteristics using an updated protocol (Lewis et al., 2018) that provides summary ratings across nine psychometric properties (see list in Methods).

## Method

### Design overview

The systematic literature review and synthesis consisted of three phases, the rationale for which is described in a previous publication (Lewis et al., 2018). Phase I, measure identification, included five steps: (1) search string generation, (2) title and abstract screening, (3) full-text review, (4) measure assignment to implementation construct(s), and (5) citation searches. Phase II, data extraction, included coding relevant psychometric information, and in Phase III data analysis was conducted.

### Phase I: data collection

First, literature searches were completed in PubMed and Embase bibliographic databases using search strings generated in consultation from PubMed support specialists and a library scientist. Consistent with our funding source and aim to identify and assess implementation-related measures in mental and behavioral health, our search was built on four core levels: (1) terms for implementation; (2) terms for measurement; (3) terms for evidence-based practice; and (4) terms

for behavioral health (i.e., mental health and substance use; Lewis et al., 2018). For the current study, we included a fifth level for each of the following *Intervention Characteristics* from the CFIR (Damschroder et al., 2009) and their associated synonyms (see Supplemental Appendix Table A1): (1) *relative advantage*, (2) *complexity*, (3) *trialability*, (4) *intervention source*, (5) *design quality and packaging*, (6) *evidence strength and quality*, (7) *adaptability*, and (8) *cost*. Literature searches were conducted independently for each aspect of Intervention Characteristics, totaling eight different sets of search strings. Articles published from 1985 to 2017 were included in the search. Searches were completed from April 2017 to May 2017 (Supplemental Appendix Table A1). Identified articles were subjected to a title and abstract screening followed by full-text review to confirm inclusion. We included only empirical studies that contained one or more quantitative measures of any of the eight intervention characteristics if they were used in an evaluation of an implementation effort in a behavioral health context.

Included articles then moved to the fourth step, construct assignment. Trained research specialists mapped measures and/or their scales to one or more of the eight aforementioned intervention characteristics (Proctor et al., 2011). Construct assignment was based on the study of author's definition of what was being measured; however, assignment was also based on item-level content coding by the research team consistent with the approach of et al. (Chaudoir et al., 2013). That is, we assigned a measure to a construct if two or more items reflected one of the eight constructs under consideration. Construct assignment was checked and confirmed by a content expert having reviewed items within each measure and/or scale.

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

### Phase II: data extraction

Once all relevant literature was identified, articles were compiled into "measure packets," that included the measure itself, the measurement development article(s) (or article with the first empirical use in a behavioral health context), and all additional empirical uses of the measure in behavioral health. To identify all relevant reports of psychometric information, the team of trained research specialists reviewed each article and electronically extracted information to assess the psychometric rating criteria recently established by the larger team, referred to hereafter as PAPERS (Psychometric And Pragmatic Evidence Rating Scale). The full rating system and criteria for the PAPERS are published elsewhere (Lewis et al., 2018; Stanick et al., 2021). The current study, which focuses on psychometric properties only, used nine relevant PAPERS criteria: (1) internal consistency, (2) convergent validity,

(3) discriminant validity, (4) known-groups validity, (5) predictive validity, (6) concurrent validity, (7) structural validity, (8) responsiveness, and (9) norms. Data on each psychometric criterion were extracted for both full measure and individual scales as appropriate. Measures were considered “unsuitable for rating” if the measure was not designed to produce psychometrically relevant information (e.g., qualitative nomination form).

Having extracted all data related to psychometric properties, the quality of information for each of the nine criteria was rated using the following scale: “poor” (−1), “none” (0), “minimal/emerging” (1), “adequate” (2), “good” (3), or “excellent” (4). Raters were two post-baccalaureate research specialists who have been embedded in the larger research team for 5+ years, working full-time on this series of systematic reviews for 2+ of those years. Rater training included coursework, as well as review of articles and books on psychometric properties and measure development. In addition, individual and group meetings twice a week with project leads who themselves are implementation measurement experts in which individual ratings were reviewed, compared, and discussed. Any point of controversy during the rating for this study was resolved by the project leads and utilized as a teaching opportunity that resulted in a manual to accompany PAPERS that has supported two new research teams in using our protocols (P. Allen et al., 2020; Khadjesari et al., 2020). Final ratings were determined from either a single score or a “rolled up median” approach, which is more reflective of the range of measure performance as compared to other summary approaches often in use (e.g., top score, worst score counts). Moreover, this approach offers potential users of a measure a single rating capturing the evidence across all previous uses that is less negative than would be produced with the “worst score counts” approach that accounts for only a single study (Terwee et al., 2012). If a measure was unidimensional or the measure had only one rating for a criterion in an article packet, then this value was used as the final rating and no further calculations were conducted. If a measure had multiple ratings for a criterion across several articles in a packet, we calculated the median score across articles to generate the final rating for that measure on that criterion. For example, if a measure was used in four different studies, each of which rated internal consistency, we calculated the median score across all four articles to determine the final rating of internal consistency for that measure. This process was conducted for each criterion.

If a measure contained a subset of scales relevant to a construct, the ratings for those individual scales were “rolled up” by calculating the median which was then assigned as the final aggregate rating for the whole measure. For example, if a measure had four scales relevant to *complexity* and each was rated for internal consistency, the median of those ratings was calculated and assigned as the final rating of internal consistency for that whole measure. This process was carried out for each psychometric criterion. When reporting the “rolled up median” approach, if the computed median resulted in a

non-integer rating, the non-integer was rounded down (e.g., internal consistency ratings of 2 and 3 would result in a 2.5 median which was rounded down to 2). In cases where the median of two scores would equal “0” (e.g., a score of −1 and 1), the lower score would be taken (e.g., −1).

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

### Phase III: data analysis

Simple statistics (i.e., frequencies) were calculated to report on measure characteristics and availability of psychometric-relevant data. A total score was calculated for each measure by summing the scores given to each of the nine psychometric criteria. The maximum possible rating for a measure was 36 (i.e., each criterion rated 4) and the minimum was −9 (i.e., each criterion rated −1). Bar charts were generated to display visual comparisons across all measures within a given construct. Following the rolled-up approach applied in this study, results are presented at the full measure level. Where appropriate, we indicate the number of scales relevant to a construct within that measure.

## Results

### Overview of measures

Searches of electronic databases yielded 16 measures related to the eight intervention characteristics constructs (*intervention source, evidence strength & quality, relative advantage, adaptability, trialability, complexity, design quality & packaging, cost*) that have been used in mental or behavioral health care research (see Figures A1–A8). No measures of *intervention source* were identified. One measure of *evidence strength & quality* was identified. Nine measures of *relative advantage* were identified, of which eight were scales within a broader measure (e.g., The Perceived Characteristics of Intervention Scale—Relative Advantage Scale; Cook et al., 2015; Moore & Benbasat, 1991). Five measures of *adaptability* were identified, four of which were scales within broader measures (e.g., Texas Christian University Organizational Readiness for Change—Adaptability Scale; Lehman et al., 2002; McLean, 2013). Six measures of *trialability* were identified, all of which were scales of broader measures (e.g., Pelzer SBI Implementation Attitude Questionnaire—Trialability Scale) (McLean, 2013; Peltzer et al., 2008). Nine measures of *complexity* were identified, all of which were scales within broader measures (e.g., Readiness for Integrated Care Questionnaire—Complexity

**Table 2.** Description of measures and subscales.

	Intervention source (N = 0)	Evidence strength & quality (N = 1)	Relative advantage (N = 9)	Adaptability (N = 5)	Trialability (N = 6)	Complexity (N = 9)	Design quality & packaging (N = 2)
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Concept defined							
Yes	0 (0)	1 (100)	4 (44)	3 (6)	3 (50)	4 (44)	1 (50)
One-Time Use Only							
Yes	0 (0)	0 (0)	2 (22)	2 (40)	4 (67)	7 (78)	2 (100)
Number of Items							
1 to 5	0 (0)	0 (0)	4 (44)	4 (80)	3 (50)	5 (56)	1 (50)
6 to 10	0 (0)	0 (0)	1 (11)	0 (0)	0 (0)	1 (11)	1 (50)
11 or more	0 (0)	1 (100)	1 (11)	1 (20)	1 (17)	0 (0)	0 (0)
Not specified	0 (0)	0 (0)	3 (33)	0 (0)	1 (17)	3 (33)	0 (0)
Country							
US	0 (0)	1 (100)	8 (88)	5 (100)	4 (67)	7 (78)	1 (50)
Other	0 (0)	0 (0)	1 (11)	0 (0)	2 (33)	2 (22)	1 (50)
Setting							
Inpatient psychiatry	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Outpatient community	0 (0)	0 (0)	3 (33)	2 (40)	2 (33)	3 (33)	0 (0)
School mental health	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Residential care	0 (0)	0 (0)	2 (22)	3 (60)	2 (33)	2 (22)	0 (0)
Other	0 (0)	1 (100)	4 (44)	2 (40)	2 (33)	4 (44)	0 (0)
Not specified	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (50)
Level							
Organization	0 (0)	0 (0)	1 (11)	0 (0)	1 (17)	1 (11)	1 (50)
Clinic/site	0 (0)	0 (0)	0 (0)	2 (40)	1 (17)	1 (11)	0 (0)
Provider	0 (0)	0 (0)	8 (89)	2 (40)	4 (67)	7 (78)	1 (50)
Director	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Supervisor	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Other	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Population							
General mental health	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Anxiety	0 (0)	0 (0)	1 (11)	1 (20)	1 (17)	1 (11)	0 (0)
Depression	0 (0)	0 (0)	1 (11)	1 (20)	1 (17)	1 (11)	0 (0)
Alcohol use disorder	0 (0)	0 (0)	0 (0)	1 (20)	1 (17)	1 (11)	0 (0)
Substance use disorder	0 (0)	0 (0)	3 (33)	3 (60)	1 (17)	4 (44)	1 (50)
Trauma	0 (0)	0 (0)	2 (22)	1 (20)	1 (17)	1 (11)	0 (0)
Not specified	0 (0)	1 (100)	0 (0)	1 (20)	2 (33)	3 (33)	1 (50)
EPIS							
Implementation	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Sustainment	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)
Not specified	0 (0)	1 (100)	9 (11)	3 (60)	6 (100)	9 (100)	2 (100)

EPIS: Exploration, Preparation, Implementation, Sustainment.

Note: No measures were tested in state mental health; No measures were tested at the consumer, system, or team level; No measures were tested in populations assessing suicidal ideation, behavioral disorders, or mania; No measures were tested for the exploration, or preparation EPIS phases.

Scale; McLean, 2013; Scott et al., 2017). Finally, two measures of *design quality & packaging* were identified, both of which were scales within broader measures (e.g., Malte Post-Treatment Smoking Cessation Beliefs Measure—Design Quality; Malte et al., 2013; McLean, 2013). Two measures of implementation *cost* were also identified; however, neither of them was suitable for rating and thus their psychometric

evidence was not assessed. It is worth noting that the number of measures listed above for each construct does not add up to 16. This is because there were nine measures identified that had scales relevant in multiple different constructs. Of these nine measures, three were included in four constructs, three were included in three constructs, and three were included in two constructs.



**Table 3.** Data availability table.

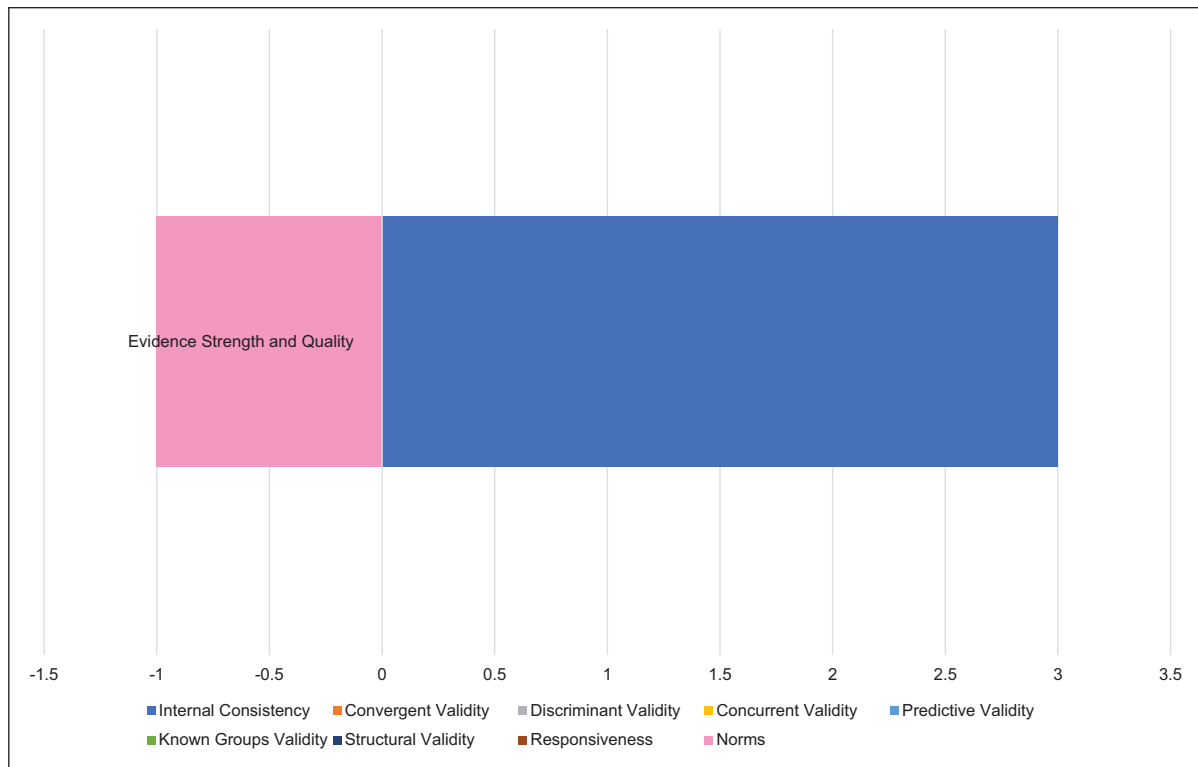
Construct	N	Internal consistency		Convergent validity		Discriminant validity		Concurrent validity		Predictive validity		Known-groups validity		Structural validity		Responsiveness		Norms	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Intervention source	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Evidence strength & quality	1	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	100
Relative advantage	9	45	-	-	-	-	-	2	23	-	-	-	-	1	12	-	-	5	56
Adaptability	5	4	80	1	20	-	-	2	40	2	40	2	40	3	60	1	20	2	40
Triability	6	3	50	-	-	-	-	2	33	-	-	-	-	1	17	-	-	2	33
Complexity	9	4	45	-	-	-	-	2	23	-	-	-	-	1	12	-	-	4	45
Design quality	2	1	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Note: Excluding measures that were deemed unsuitable for rating.

**Table 4.** Psychometric summary data.

Construct	N	Internal consistency		Convergent validity		Discriminant validity		Concurrent validity		Predictive validity		Known-groups validity		Structural validity		Responsiveness		Norms	
		M	(R)	M	(R)	M	(R)	M	(R)	M	(R)	M	(R)	M	(R)	M	(R)	M	(R)
Intervention source	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Evidence strength & quality	1	3	(3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1	(-1)
Relative advantage	9	2	(1-4)	-	-	-	-	-	-	2	(1-4)	-	-	3	(3)	-	-	1	(-1 to 2)
Adaptability	5	2	(1-4)	-1	(-1 to 2)	-	-	-	-	-1	(-1 to 1)	3	(2-3)	2	(2-3)	3	(3)	2	(2)
Triability	6	2	(1-3)	-	-	-	-	-	-	2	(1-4)	-	-	3	(3)	-	-	-1	(-1 to 2)
Complexity	9	2	(2-3)	-	-	-	-	-	-	3	(1-4)	-	-	3	(3)	-	-	-1	(-1 to 2)
Design quality	2	3	(3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1	(-1 to 2)

Note: M: median; R: range; excluding zeros where psychometric information was not available; excluding measures that were deemed unsuitable for rating; possible scores ranged from “-1” (poor) to “4” (excellent).



**Figure 1.** Evidence strength and quality ratings.

### Characteristics of measures

Table 2 presents the descriptive characteristics of measures used to assess constructs of intervention characteristics. Most measures of intervention characteristics were created in the United States ( $n = 13$ , 81%). The remaining measures were developed in Canada and South Africa. Half of the measures were either used in residential care settings ( $n = 3$ , 19%) or outpatient care settings ( $n = 5$ , 31%) and the rest were used in a variety of “other” settings (e.g., primary care, military, university;  $n = 7$ , 44%). Half of the measures were used to assess intervention characteristics influencing implementation in alcohol or substance use treatment ( $n = 7$ , 44%) and with the rest assessing treatment for trauma ( $n = 2$ , 13%), or a target problem was not specified.

### Availability of psychometric evidence

Of the 16 measures and scales of implementation constructs, two cost formulas were categorized as unsuitable for rating per our protocol. For the remaining 14 measures, there was limited psychometric information available (Table 3). Seven (50%) measures had no information for internal consistency, 13 (93%) had no information for convergent validity, no measures (100%) had information for discriminant validity or for concurrent validity, 11 (79%) had no information for predictive validity, 12 (86%) had no information for known-groups

validity, 11 (79%) had no information for structural validity, 13 (93%) had no information for responsiveness, and finally, 7 (50%) had no information on norms.

### Psychometric Evidence Rating Scale results

Table 4 provides the median ratings and range of ratings for psychometric properties for measures deemed suitable for rating ( $n = 14$ ) and those for which information was available (i.e., those with non-zero ratings on PAPERS criteria). Individual ratings for all measures can be found in Supplemental Appendix Table A2 and head-to-head bar graphs can be found in Figures 1 to 6.

**Evidence strength & quality.** One measure of *evidence strength & quality* was identified in mental or behavioral health care research. Information about internal consistency and norms was available with no information available for the other psychometric properties. The rating for internal consistency was “3—good” and “-1—poor” for norms with a total combined score of “two—adequate” (maximum possible score = 36; Barriers to Treatment Integrity Implementation Survey; Perepletchikova et al., 2009).

**Relative advantage.** Nine measures of *relative advantage* were identified in mental or behavioral health care research. Information about internal consistency was available for four measures (45%), convergent validity for

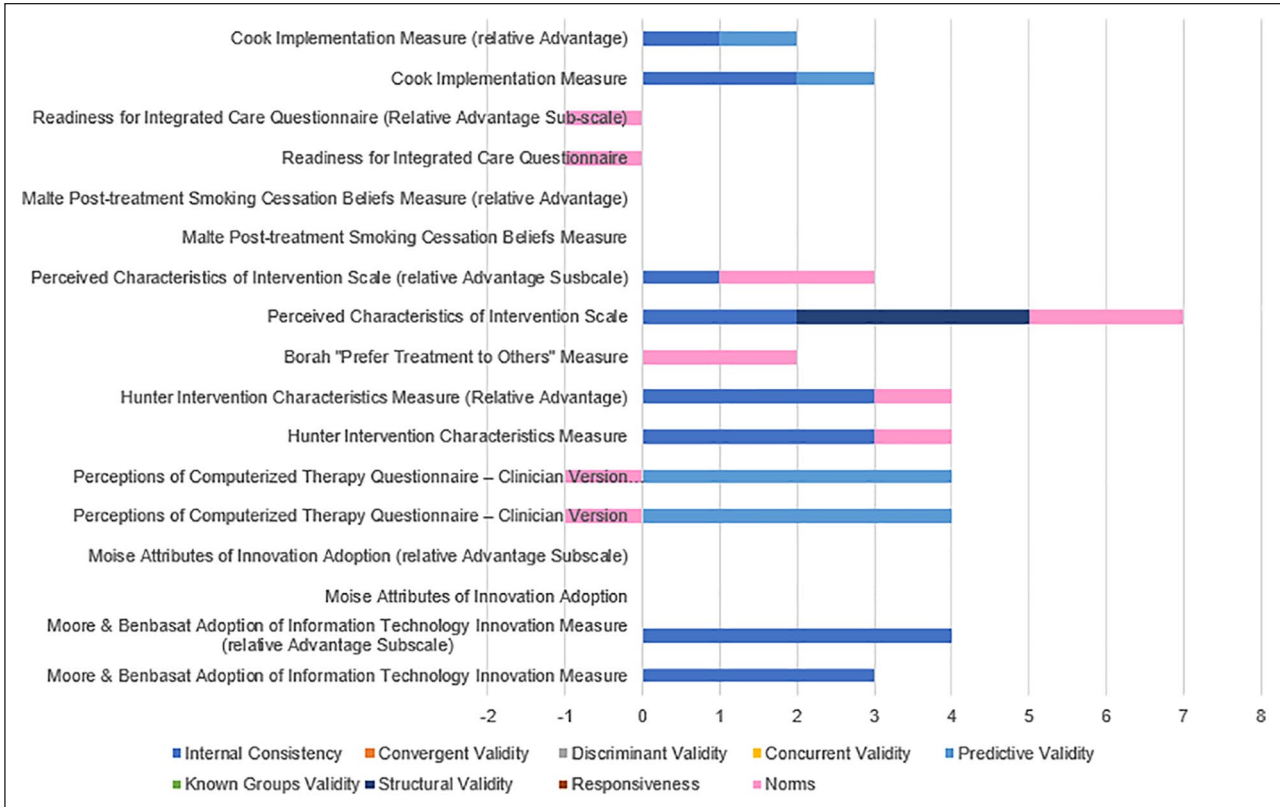


Figure 2. Relative advantage ratings.

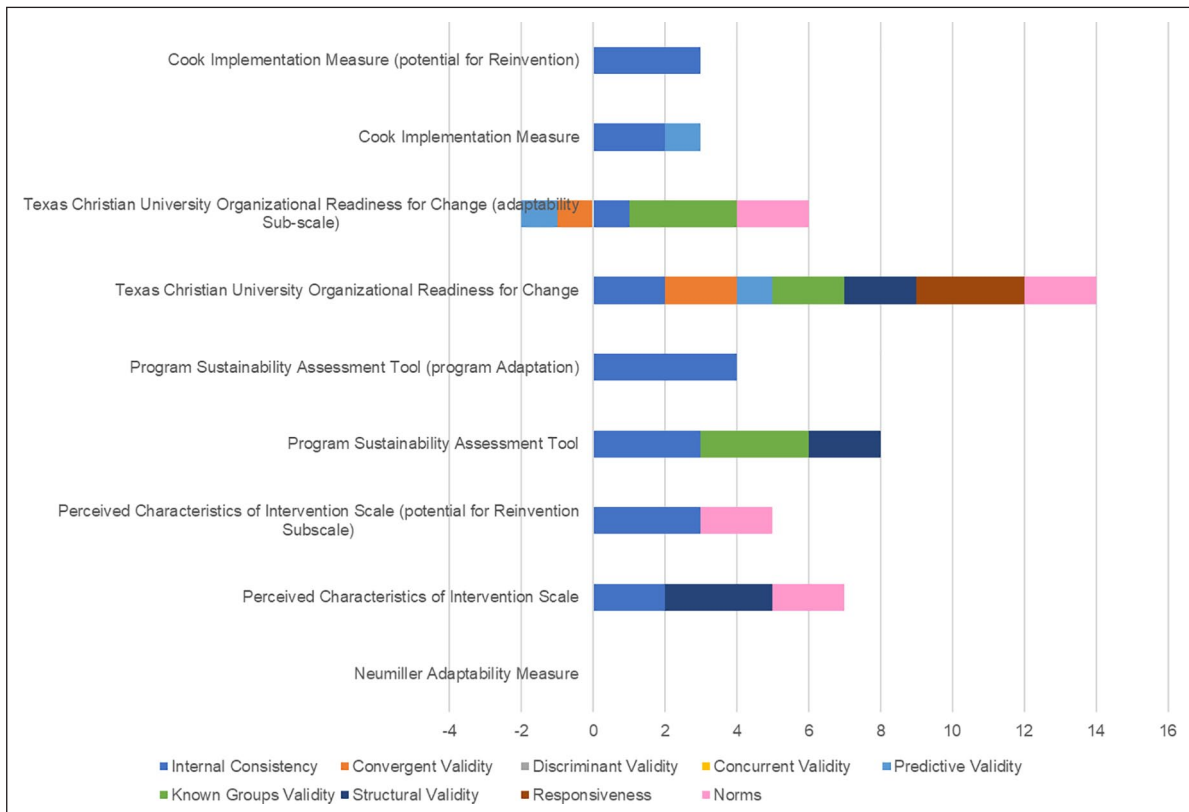


Figure 3. Adaptability ratings.



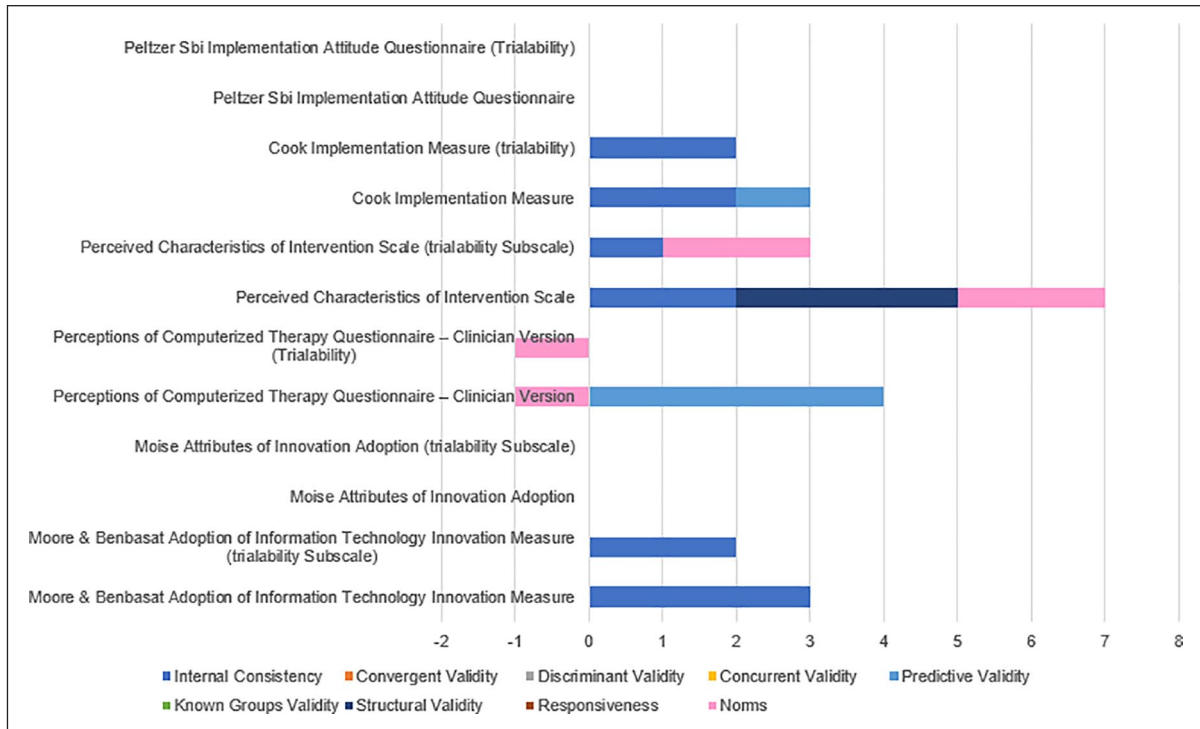


Figure 4. Trialability ratings.

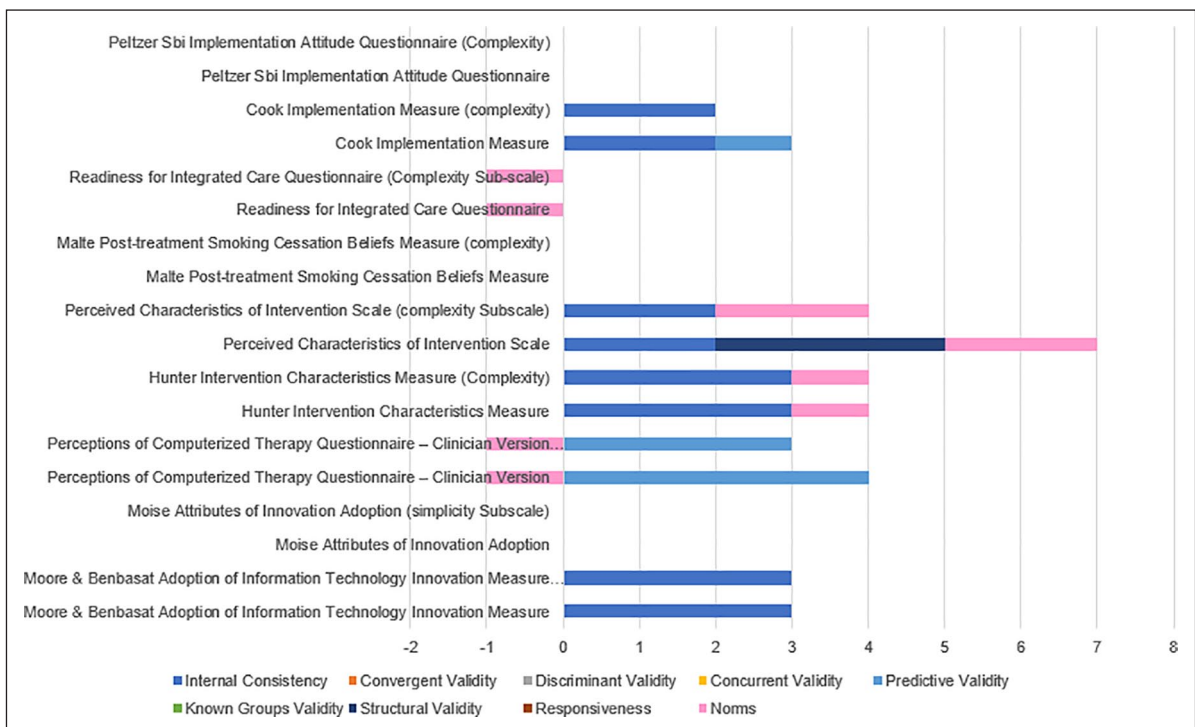
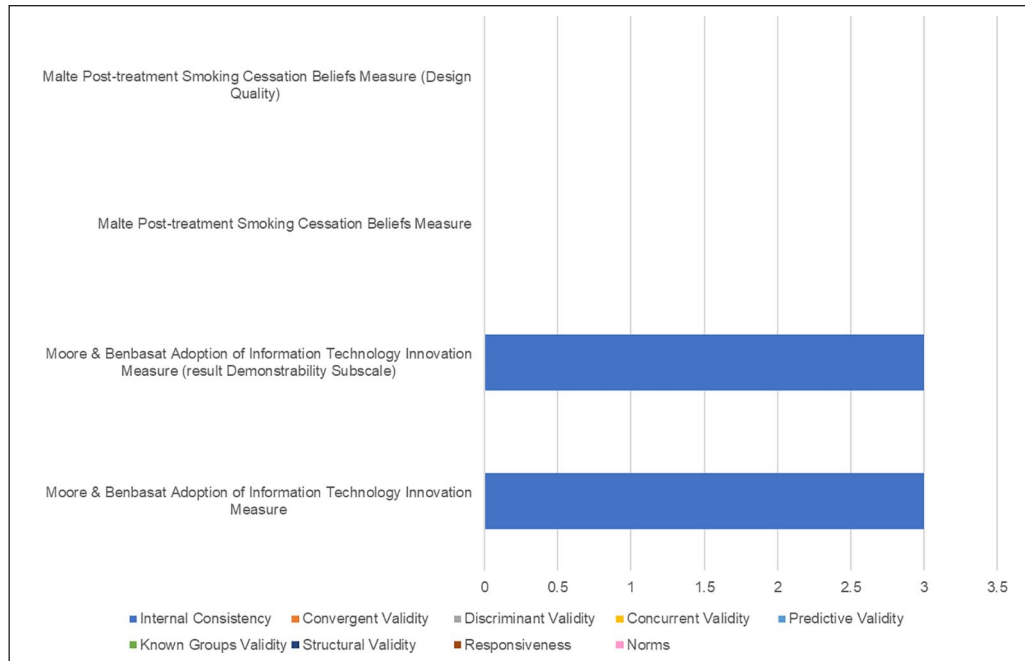


Figure 5. Complexity ratings.

no measures, discriminant validity for no measures, concurrent validity for no measures, predictive validity for two measures (23%), known-groups validity for no

measures, structural validity for one measure (12%), responsiveness for no measures, and norms for five measures (56%). For those measures of *relative advantage* with



**Figure 6.** Design quality and packaging ratings.

information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,” “2—adequate” for predictive validity, “3—good” for structural validity, “1—minimal/emerging” for norms.

The Hunter Intervention Characteristics Relative Advantage scale and Moore and Benbasat Adoption of Information Technology Innovation Relative Advantage scale had the highest psychometric rating scores among measures of *relative advantage* used in mental and behavioral health care (psychometric total maximum scores = [4]; maximum possible score = 36), with ratings of “3—good” and “4—excellent” respectfully for internal consistency and “1—minimal/emerging” for norms in Hunter and colleagues measure (Hunter et al., 2015; Moore & Benbasat, 1991; Yetter, 2010). There was no information available on any of the remaining psychometric criteria.

**Adaptability.** Five measures of *adaptability* were identified in mental or behavioral health care research. Information about internal consistency was available for four measures (80%), convergent validity for one measure (20%), discriminant validity for no measures, concurrent validity for no measures, predictive validity for two measures (40%), known-groups validity for two measures (40%), structural validity for three measures (60%), responsiveness for one measure (20%), and norms for two measures (40%). For those measures of *adaptability* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,”

“-1—poor” for convergent validity, “-1—poor” for predictive validity, “3—good” for known-groups validity, “2—adequate” for structural validity, “3—good” for responsiveness, and “2—adequate” for norms.

The Perceived Characteristics of Intervention Scale Potential for Reinvention scale had the highest psychometric rating score among measures of *adaptability* used in mental and behavioral health care (psychometric total maximum score = 5; maximum possible score = 36), with ratings of “3—good” for internal consistency and “2—adequate” for norms (Cook et al., 2015). There was no information available on any of the remaining psychometric criteria.

**Trialability.** Six measures of *trialability* were identified in mental or behavioral health care research. Information about internal consistency was available for three measures (50%), convergent validity for no measures, discriminant validity for no measures, concurrent validity for no measures, predictive validity for two measures (34%), known-groups validity for no measures, structural validity for one measure (17%), responsiveness for no measures, and norms for two measures (34%). For those measures of *trialability* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,” “2—adequate” for predictive validity, “3—good” for structural validity, and “-1—poor” for norms.

As with *relative advantage*, the Perceived Characteristics of Intervention Trialability scale had the highest psychometric rating score among measures of *trialability* used in

mental and behavioral health care (psychometric total maximum score = [3]; maximum possible score = 36), with ratings of “1—minimal/emerging” for internal consistency and “2—adequate” for norms (Cook et al., 2015; Yetter, 2010). There was no information available on any of the remaining psychometric criteria.

**Complexity.** Nine measures of *complexity* were identified in mental or behavioral health care research. Information about internal consistency was available for four measures (45%), convergent validity for no measures, discriminant validity for no measures, concurrent validity for no measures, predictive validity for two measures (23%), known-groups validity for no measures, structural validity for one measure (12%), responsiveness for no measures, and norms for four measures (45%). For those measures of *complexity* with information available (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was “2—adequate,” “3—good” for predictive validity, “3—good” for structural validity, and “1—poor” for norms.

The Hunter Intervention Characteristics Relative Advantage scale and the Perceived Characteristics of Intervention Complexity scale had the highest psychometric rating scores among measures of *complexity* used in mental and behavioral health care (psychometric total maximum score = [4]; maximum possible score = 36), with ratings of “3—good” and “2—adequate” respectfully for internal consistency and “1—minimal/emerging” and “2—adequate” respectfully for norms (Cook et al., 2015; Hunter et al., 2015; Yetter, 2010). There was no information available on any of the remaining psychometric criteria.

**Design quality & packaging.** Two measures of *design quality & packaging* were identified in mental or behavioral health care research. Information about internal consistency was available for both instruments, with no information available for the other psychometric properties. The rating for internal consistency was “3—good” (Adoption of Information Technology Innovation Result Demonstrability scale; Moore & Benbasat, 1991).

## Discussion

### Summary of study findings

This systematic review identified 16 unique measures and/or subscales across the eight constructs in the domain of intervention characteristics, several of which were assigned to multiple constructs: zero for *intervention source*, one for *evidence strength and quality*, nine for *relative advantage*, five for *adaptability*, six for *trialability*, nine for *complexity*, and two for *design quality and packaging*. Of the identified measures and scales, only about 50% defined the constructs they endeavored to assess, making it difficult to

have confidence in these measures given threats to content validity. Nearly all measures were developed in the United States, with the remaining three measures coming from Canada and South America, suggesting a bolus of measure development work in this understudied domain coming from a single region. Half of the measures had been used in studies targeting substance use and nearly all targeted the provider, which may indicate that researchers in this area are particularly sensitive to how perceptions of the intervention influence successful implementation perhaps given that substance use is often a comorbid condition demanding more complex interventions. Finally, the majority had only ever been used once, which reflects the nascent state of exploring this “forgotten domain” of implementation determinants.

Consistent with findings from our team’s other systematic reviews across the CFIR domains (Lewis et al., 2016), the most common psychometric properties reported across all identified measures were internal consistency and norms. This is most likely the case because many journals have minimal reporting standards for measures that require statistical values (i.e., alpha, *M*, *SD*) for these properties be provided for a given sample. It is promising to see the potential impact of reporting standards on building the evidence base for implementation measures.

Unfortunately, that is where measurement reporting standards ends. Consequently, no measures had any data contributing evidence of discriminant validity or concurrent validity, and only one measure had any evidence of convergent validity. The limited number of measures and their associated psychometric data confirms that few scholars are exploring the role of intervention characteristics in behavioral health implementation, particularly as they relate to other constructs or established measures, calling into question the construct validity of the identified measures. Moreover, only one measure had any evidence of responsiveness, which indicates that exploring how implementation strategies affect perceptions of intervention characteristics is rarely conducted. Finally, when available, psychometric evidence across criteria ranged from poor to good, with only three instances across the numerous applications listed in Table 4 reflective of “excellent” evidence. These results suggest a significant amount of work is needed to build the quality of measures of intervention characteristics.

Across the constructs, measures of *adaptability* had been subjected to the most testing yielding promising psychometric evidence across the following properties among the identified measures: internal consistency, convergent validity, predictive validity, known-groups validity, structural validity, responsiveness, and norms. This may be because some tout that there is “no implementation without adaptation” (Lyon & Bruns, 2019, p. 3) thus placing demands on interventions to be adaptable as they enter new service delivery settings. Beyond instruments that

measure adaptability, the growing focus on intervention adaptation in implementation more generally (e.g., Chambers & Norton, 2016; Stirman et al., 2019) indicates that instruments assessing all manner of intervention characteristics may be useful to evaluate the outcomes of adaptation efforts. For instance, intervention adaptation may be undertaken with the explicit objective of improving *design quality* (Lyon et al., 2019).

We identified zero measures of *intervention source* suggesting few researchers perceive this construct to be of significance compared to others in this domain. Although we identified two measures of *design quality and packaging*, no properties, other than internal consistency, offered evidence of psychometric quality. To answer the questions laid out in the introduction about the differential influence of intervention characteristics over the course of implementation, more research is needed across all constructs in this domain. Concentrated work on this domain is warranted, particularly when compared with the sheer number of measures identified for single constructs at other system levels like “leadership engagement” (inner setting level) for which 24 unique measures emerged (Weiner et al., 2020).

### Comparison with previous systematic reviews

The results of our systematic review align with findings of two previous reviews stating that information of measures’ psychometric properties is often scarce or underreported. Previous reviews found that only 38% of identified measures demonstrated any evidence of reliability and validity. In our review, 64% of measures identified showed evidence for at least one of the nine psychometric properties assessed. However, most evidence reported was only minimal or emerging with none scoring above a 10 on our PAPER scale (total possible score of 36; Stanick et al., 2021). It is also worth noting the discrepancy in the number of measures. Across the two previous reviews, they identified 43 unique measures (27 in Chor et al., 2015; 16 in Chaudoir et al., 2013), whereas our review only identified 16 measures of intervention characteristics. After careful examination by our authorship team of differences between our study protocols, it is the case that the limited number of measures we identified is due to our focus on those used in behavioral health settings.

### Limitations

There are several noteworthy limitations of this systematic review. First, our focus on behavioral health means that we overlooked several measures of these constructs that have yet to be used in these settings. Another consequence of our focus is that identified measures with uses in other settings did not contribute to the evidence synthesis reported here. In comparing our findings with those of the two previously published systematic reviews, it seems the *intervention characteristics* domain may be most vulnerable to single

use measures given that scholars tend to develop them with such specificity to the intervention that limits their applicability for broad use. This may especially be the case for measures of *intervention characteristics* developed for complex psychosocial interventions used in behavioral health. Second, our systematic search was completed in 2017, but the citation searches were completed in 2019. Because of this, we may have missed new measures developed in the past 3 years, but we are confident that we have captured the most recent uses of identified measures in this report. Third, poor reporting practices limit the extent to which evidence was available for the identified measures. It is possible that more testing was conducted but simply not reported, or was reported in a way that did not allow for our nuanced rating system to detect variations in quality (e.g.,  $\alpha > .70$ , rather than stating the exact values). Fourth, some of the identified measures are likely specific to a given intervention (e.g., digital technology) that may limit the generalizable use of the measure. Fifth, our rolled up median approach for summarizing psychometric evidence across studies does not precisely reflect measures’ best or worst performance and so results should be interpreted with this in mind. Sixth, it is possible that, despite our team’s rigorous approach to construct assignment, that some measures were misaligned and so we encourage users and measure developers to carefully review item level content for themselves. Finally, we were unable to apply the pragmatic measures rating criteria in this systematic review as it was developed contemporaneously and not available to our study team; this may limit the utility of the results.

### Implications and future directions

Given the extensiveness of the information presented, the utility of different aspects of this paper may vary by professional role. Implementation practitioners interested in systematically exploring intervention characteristics are encouraged to consume the figures included in this manuscript. Overall bar length may be primary in deciding which measure to select within a given construct. Or, if the intended purpose of the measure data is clear, then certain colors/segments of each bar may be more discerning. Measure developers may find the data availability (Table 3) most useful if they are trying to fill a gap, or the psychometrics summaries (Table 4) if they are trying to build on a promising measure. Implementation practitioners need pragmatic measures that are reliable and valid, but they also need more evidence from researchers about which intervention characteristics matter when and why in implementation. Until the empirical gaps raised in the introduction are filled, it is likely that intervention developers will continue to develop complex, costly interventions that do not offer relative advantage over existing interventions, are difficult to try, and are packaged in ways that do not promote adoption or adaptation. These empirical gaps are the foundation of observed implementation gaps, and ulti-



mately addressing these empirical gaps require psychometrically strong measures.

Finally, from a measurement perspective, it is critical to bring clarity to study participants on the referent, which in this case is the intervention itself. In the context of behavioral health, the intervention is often a complex psychosocial intervention that may include multiple components over time delivered by one or more individuals. Exposing participants (patients/clients, providers, leaders, etc.) to the full scope of the intervention so that they can respond accurately about its relative advantage, complexity, and trialability, for instance, remains a challenge. Accordingly, reliable and valid measurement of intervention characteristics may require creative ways to articulate the intervention itself in the instructions of a measure.

## Conclusion

High-quality assessment instruments provide a critical foundation for conducting implementation research. The current study confirms that intervention characteristics are understudied in behavioral health implementation research, at least with regard to the development and testing of measures intended to assess “the thing” being implemented (Curran, 2020). As the emphasis on intervention characteristics (e.g., *adaptability*, *design quality*) continues to grow, measurement approaches that keep pace with and supply the research and practice community with rigorous tools to understand aspects of innovations that influence—and are influenced by—implementation efforts will be sorely needed.

## Declaration of conflicting interests

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

## Funding

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

## ORCID iDs

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

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

Aaron R Lyon  <https://orcid.org/0000-0003-3657-5060>

## Supplemental material

Supplemental material for this article is available online.

## References

- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4–23. <https://doi.org/10.1007/s10488-010-0327-7>
- Allen, J. D., Towne, S. D., Maxwell, A. E., DiMartino, L., Leyva, B., Bowen, D. J., & Weiner, B. J. (2017). Measures of organizational characteristics associated with adoption and/or implementation of innovations: A systematic review. *BMC Health Services Research, 17*(1), Article 591. <https://doi.org/10.1186/s12913-017-2459-x>
- Allen, P., Pilar, M., Walsh-Bailey, C., Hooley, C., Mazzucca, S., Lewis, C. C., Mettert, K. D., Dorsey, C. N., Purtle, J., Kepper, M. M., Baumann, A. A., & Brownson, R. C. (2020). Quantitative measures of health policy implementation determinants and outcomes: A systematic review. *Implementation Science, 15*(1), 47. <https://doi.org/10.1186/s13012-020-01007-w>
- Borah, E. V., Wright, E. C., Donahue, D. A., Cedillos, E. M., Riggs, D. S., Isler, W. C., & Peterson, A. L. (2013). Implementation outcomes of military provider training in cognitive processing therapy and prolonged exposure therapy for post-traumatic stress disorder. *Military Medicine, 178*(9), 939–944. <https://doi.org/10.7205/MILMED-D-13-00072>
- Carper, M. M., McHugh, R. K., & Barlow, D. H. (2013). The dissemination of computer-based psychological treatment: A preliminary analysis of patient and clinician perceptions. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(2), 87–95. <https://doi.org/10.1007/s10488-011-0377-5>
- Chambers, D. A., & Norton, W. E. (2016). The adaptome: Advancing the science of intervention adaptation. *American Journal of Preventive Medicine, 51*(4, Suppl. 2), S124–S131. <https://doi.org/10.1016/j.amepre.2016.05.011>
- 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, 42*(5), 545–573. <https://doi.org/10.1007/s10488-014-0551-7>
- 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>
- Cook, J. M., Thompson, R., & Schnurr, P. P. (2015). Perceived characteristics of intervention scale: Development and psychometric properties. *Assessment, 22*(6), 704–714. <https://doi.org/10.1177/1073191114561254>



- Curran, G. M. (2020). Implementation science made too simple: A teaching tool. *Implementation Science Communications, 1*(1), 27. <https://doi.org/10.1186/s43058-020-00001-z>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. <https://doi.org/10.1186/1748-5908-4-50>
- Hunter, S. B., Han, B., Slaughter, M. E., Godley, S. H., & Garner, B. R. (2015). Associations between implementation characteristics and evidence-based practice sustainment: A study of the Adolescent Community Reinforcement Approach. *Implementation Science, 10*, 173. <https://doi.org/10.1186/s13012-015-0364-4>
- Khadjesari, Z., Boufkhed, S., Vitoratou, S., Schatte, L., Ziemann, A., Daskalopoulou, C., Ugluk-Marucha, E., Sevdalis, N., & Hull, L. (2020). Implementation outcome instruments for use in physical healthcare settings: A systematic review. *Implementation Science, 15*(1), 66. <https://doi.org/10.1186/s13012-020-01027-6>
- Lehman, W. E., 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., Stanick, C. F., Martinez, R. G., Weiner B. J., Kim, M., Barwick, M., & Comtois, K. A. (2016). The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. *Implementation Science 10*(2), <https://doi.org/10.1186/s13012-014-0193-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*(1), 66. <https://doi.org/10.1186/s13643-018-0728-3>
- 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>
- Lyon, A. R., & Bruns, E. J. (2019). User-centered redesign of evidence-based psychosocial interventions to enhance implementation-hospitable soil or better seeds? *Journal of the American Medical Association Psychiatry, 76*(1), 3–4. <https://doi.org/10.1001/jamapsychiatry.2018.3060>
- Lyon, A. R., Munson, S. A., Renn, B. N., Atkins, D. C., Pullmann, M. D., Friedman, E., & Areal, P. A. (2019). Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: Protocol for studies applying a framework to assess usability. *Journal of Medical Internet Research, Research Protocols, 8*(10), Article e14990. <https://doi.org/10.2196/14990>
- Malte, C. A., McFall, M., Chow, B., Beckham, J. C., Carmody, T. P., & Saxon, A. J. (2013). Survey of providers' attitudes toward integrating smoking cessation treatment into posttraumatic stress disorder care. *Psychology of Addictive Behaviors, 27*(1), 249–255. <https://doi.org/10.1037/a0028484>
- McLean, K. A. (2013). *Healthcare provider acceptability of a behavioral intervention to promote adherence* [Open Access Theses, University of Miami]. [https://scholarlyrepository.miami.edu/oa\\_theses/434](https://scholarlyrepository.miami.edu/oa_theses/434)
- Miake-Lye, I. M., Delevan, D. M., Ganz, D. A., Mittman, B. S., & Finley, E. P. (2020). Unpacking organizational readiness for change: An updated systematic review and content analysis of assessments. *BMC Health Services Research, 20*(1), Article 106. <https://doi.org/10.1186/s12913-020-4926-z>
- Moise, I. K., Green, D., Toth, J., & Mulhall, P. F. (2014). Evaluation of an authority innovation-decision: Brief alcohol intervention for pregnant women receiving women, infants, and children services at two Illinois health departments. *Substance Use & Misuse, 49*(7), 804–812. <https://doi.org/10.3109/10826084.2014.880484>
- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. <https://doi.org/10.1287/isre.2.3.192>
- Moullin, J. C., Dickson, K. S., Stadnick, N. A., Rabin, B., & Aarons, G. A. (2019). Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. *Implementation Science, 14*(1), 1. <https://doi.org/10.1186/s13012-018-0842-6>
- Neumiller, S., Bennett-Clark, F., Young, M. S., Dates, B., Broner, N., Leddy, J., & De Jong, F. (2009). Implementing assertive community treatment in diverse settings for people who are homeless with co-occurring mental and addictive disorders: A series of case studies. *Journal of Dual Diagnosis, 5*(3–4), 239–263. <https://doi.org/10.1080/15504260903175973>
- Peltzer, K., Matseke, G., & Azwihangwisi, M. (2008). Evaluation of alcohol screening and brief intervention in routine practice of primary care nurses in Vhembe district, South Africa. *Croatian Medical Journal, 49*(3), 392–401. <https://doi.org/10.3325/cmj.2008.3.392>
- Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology, 77*(2), 212–218. <https://doi.org/10.1037/a0015232>
- 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, 38*(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Rabin, B. A., Lewis, C. C., Norton, W. E., Neta, G., Chambers, D., Tobin, J. N., & Brownson, R. C. (2016). Measurement resources for dissemination and implementation research in health. *Implementation Science, 11*, 42. <https://doi.org/10.1186/s13012-016-0401-y>
- 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>

- 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>
- Stirman, S. W., Baumann, A. A., & Miller, C. J. (2019). The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science, 14*(1), 58.
- 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>
- Turzhitsky, V., Rogers, J. D., Mutyal, N. N., Roy, H. K., & Backman, V. (2010). Characterization of light transport in scattering media at sub-diffusion length scales with Low-coherence Enhanced Backscattering. *Journal of Quantum Electronics, 16*(3), 619–626. <https://doi.org/10.1109/JSTQE.2009.2032666>
- Weiner, B. J., Mettert, K. D., Dorsey, C. N., Nolen, E. A., Stanick, C., Powell, B. J., Lewis, C. C. (2020). Measuring readiness for implementation: A systematic review of measures' psychometric and pragmatic properties. *Implementation Research and Practice*. Advance online publication. <https://doi.org/10.1177/2633489520933896>
- Willmeroth T, Wesselborg B, Kuske S. (2019). Implementation Outcomes and Indicators as a New Challenge in Health Services Research: A Systematic Scoping Review. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*. Advance online publication. <https://doi.org/10.1177/0046958019861257>
- Yetter, G. (2010). Assessing the acceptability of problem-solving procedures by school teams: Preliminary development of the pre-referral intervention team inventory. *Journal of Educational and Psychological Consultation, 20*(2), 139–168. <https://doi.org/10.1080/10474411003785370>