

REVIEW

Open Access



Scoping review of instruments measuring organizational context in implementation studies within healthcare

Janet E. Squires^{1,2*}, Ian D. Graham^{1,2,3}, Wilmer J. Santos^{1,2}, Ligya Korki de Candido¹, Vanessa J. Watkins⁴, Megan Greenough^{1,5}, Alison M. Hutchinson^{4,6} and the ICON Team

Abstract

Background Healthcare providers' use of research in practice is suboptimal. In Learning Health Systems, using the best available research evidence to guide practice is fundamental to continuous improvement. Whilst there is a growing body of evidence that context modifies the effects of implementation strategies to promote research use, knowledge on how best to measure context is lacking. As part of a larger research programme, which involved development of the Implementation in Context (ICON) framework comprising six domains and 22 attributes (Fig. 1), we conducted a scoping review to identify psychometrically sound and pragmatic instruments for measuring macro- or meso-level ICON attributes.

Methods The review was conducted in two phases: (1) identification of instruments that align with the 20 meso- and macro-level attributes in ICON and (2) assessment of their psychometric and pragmatic properties. Instruments and psychometric papers were identified through five electronic databases. The COnsensus-based Standards for selecting health Measurement INstruments (COSMIN) criteria guided the psychometric evaluation. A modified GRADE approach was used to summarize the quality of the evidence for each instrument. In this paper, we report the properties of the included instruments measuring one of the macro- or meso-level attributes.

Results We identified 38 instruments (and assessed 169 psychometric papers) that captured nine of 15 meso-level ICON attributes; instruments measuring macro-level ICON attributes were not located. Most of the 38 instruments were rated as inadequate or doubtful ($n=31$) on content validity. Using the COSMIN recommendation ratings, 13 instruments are recommended for measuring five meso-level ICON attributes: organizational climate, organizational change processes, receptivity to change, evaluation activity, and leadership. The remaining 25 instruments have the potential for use. All 38 instruments were rated to be at least adequate on pragmatic criteria.

Conclusions We identified 13 instruments that measure meso-level ICON attributes that influence research implementation and meet psychometric quality criteria. Future work will include: (1) additional psychometric assessments of the 25 instruments not yet recommended for use, (2) psychometric and pragmatic assessments of instruments that assess multiple attributes in ICON, and (3) the development of instruments to measure ICON attributes for which instruments were not identified.

Registration Open Science Framework, 21 June 2022, <https://doi.org/10.17605/OSF.IO/B2Z8K>

Keywords Delivery of Health Care, Implementation science, Knowledge translation, Implementation in Context (ICON) framework, Measurement, Instrument, Psychometrics, Scoping review, COSMIN

*Correspondence:

Janet E. Squires

janet.squires@uottawa.ca

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Healthcare providers' use of research in clinical practice is suboptimal, despite increased awareness of evidence and known gaps [1–9]. In a recent knowledge synthesis of inappropriate healthcare across Canada, on average, 30% of clinical practices reported in 174 studies were *not* aligned with the best available evidence [9]. Suboptimal levels of research-informed care have also been reported globally: the United States (45%) [5]; United Kingdom, Australia, and New Zealand (51%) [8]; and Australia (43%) [7]. In response, there is widespread interest in the science and practice of research implementation to advance Learning Health Systems, which aim to integrate clinical practice with research, enabling continuous learning and improvement in healthcare [10]. Researchers have identified several major methodological issues facing the implementation field [11–16], including measuring and reporting the influence of context in the design, assessment, and outcomes of implementation strategies. Contextual factors can impact the effectiveness of implementation strategies by exacerbating the problem the strategy is designed to overcome, influencing the susceptibility of the problem to the strategy, and determining how the strategy can work [17]. Context differs between settings and changes over time, making its assessment critical to the design and delivery of optimally effective implementation strategies, facilitating evaluation of the transferability of implementation strategies between settings, and optimizing the sustainability of implementation strategies [17]. Whilst there is growing recognition that context can modify the effects of implementation strategies, we lack knowledge on how best to measure context [18, 19]. Evaluating context within Learning Health Systems is crucial to identifying key questions and facilitating actions for continuous improvement. Authors of a recent realist synthesis found that effective evaluations depend on fostering supportive organizational conditions, rapid feedback loops, and a culture of innovation [20]. Understanding these contextual factors informs research implementation and drives the sustainability of health innovation. To provide conceptual clarity regarding the attributes of context, we conducted a large body of research, which led to the development of the Implementation in CONtext (ICON) framework, described in the next section. To understand how contextual attributes operate in a specific setting within a Learning Health System, implementers and researchers require valid, reliable, and feasible measures. Use of such measures, importantly, can provide an understanding of contextual attributes prior to implementation, inform strategies to address attributes identified as potential influences on implementation success, and provide an understanding of contextual attributes following implementation.

The implementation in CONtext (ICON) framework

Figure 1 depicts the ICON framework, which guided the project reported in this paper. Development of ICON was prompted by a lack of conceptual clarity about context in implementation. Thus, our earlier work focused on the first steps to gain clarity about implementation context through a series of four studies. Study 1 was a concept analysis of context in implementation ($N=70$ articles) [21, 22]. Study 2 was an analysis of 145 interviews with a diverse sample of healthcare providers, exploring elements of context relevant to their use of research findings in clinical practice [23]. Study 3 comprised interviews with 39 health system leadership stakeholders in four high-income countries to elicit their perceptions of important elements of context in implementation efforts [24]. Study 4 involved the triangulation of results from studies 1–3 to produce the ICON framework, comprising 3 levels, 6 domains, 22 core attributes, and 108 features of context relevant to implementation [25] (Fig. 1). Definitions of the 22 core attributes of ICON are located in the manuscript detailing its development [25]. More recently, we conducted a study that indicates all ICON attributes also apply in low- and middle-income countries [26]. The ICON framework provides conceptual clarity on context in implementation, which we define as: the characteristics of the provider group and the healthcare recipient group, the internal organizational characteristics (climate and structures, social behavior, response to change, and processes), and the external influences of the broader healthcare system. These contextual domains have the potential to directly or indirectly influence or modify the use of research evidence by healthcare providers, policy-makers, and recipients of care. ICON is the only implementation meta-framework solely dedicated to the role of context in implementation. While context is included in several other implementation frameworks—for example, the Promoting Action on Research Implementation in Health Services (PARIHS) Framework [27, 28], Knowledge-to-Action Framework (KTA) [29], and the Consolidated Framework for Implementation Research (CFIR) [30, 31]—it is not sufficiently detailed in these frameworks. For instance, many of the features of context attributes reported in ICON are not identified in other implementation frameworks; examples include technology (attribute *Resource Access*), mentorship (attribute *Leadership*), continuity of care (attribute *Work Structure*), funding model (attribute *Financial*), social interactions (attribute *Collaboration*), accountability (attribute *Professional Role*), and system complexity (attribute *System Features*) [32].

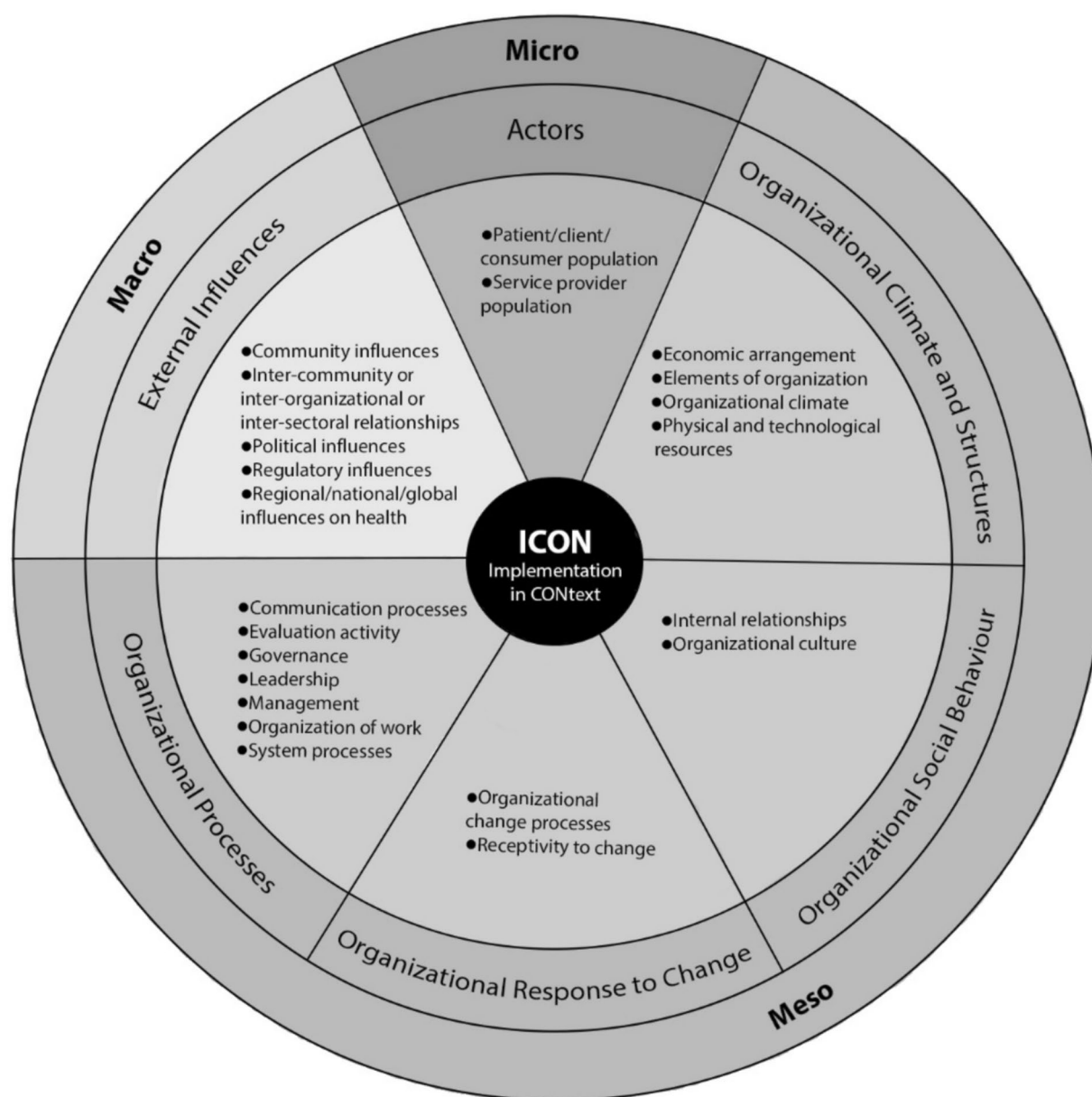


Fig. 1 The Implementation in Context (ICON) Framework. Outer circle represents levels of context. Middle circle represents the six context domains within ICON. Inner circle represents the 22 core attributes of context within each domain in ICON

Previous measurement reviews on context

A small number of syntheses [33–41] conducted by other research teams have identified context instruments, but they are based on either incomplete conceptualizations of context and/or were limited to a single setting. For example, Chaudoir and colleagues [33] assessed measures for six ICON attributes (organizational culture, physical and technological resources, economic arrangements, governance, leadership, organizational climate), limited to quality improvement studies. Brennan

et al. [34] assessed measures for seven ICON attributes (service provider population, organizational climate, organizational culture, leadership, physical and technological resources, communication processes, receptivity to change), in primary care only. Rogers et al. [35] reported a synthesis of 64 studies on context measurement; most instruments measured a single ICON attribute or were developed for specific innovations and are, therefore, not suitable for generic context assessments in implementation across innovations or healthcare sectors.

More recently, the Society for Implementation Research Collaboration (SIRC) assessed measures that map onto the CFIR framework, which categorizes and describes contextual determinants across five domains that may influence implementation [30, 31]. Their work is published in five reviews, each targeting selected constructs in four of the five context-related CFIR domains: (1) outer setting (e.g., cosmopolitanism, peer pressure) [37], (2) inner setting (e.g., leadership engagement, organizational culture) [38–40], (3) characteristics of individuals (e.g., knowledge and beliefs about the innovation, self-efficacy) [41], and (4) implementation processes (e.g., planning, engaging) [39]. Although these reviews significantly expanded the knowledge base on how to measure context, several gaps remained. First, while CFIR includes 18 constructs related to context, all but one (culture) are features (examples) of the 22 ICON context attributes. Altogether, the 18 constructs in CFIR represent only 28% (38 of 136) of the context features/constructs that comprise the 22 context attributes in ICON. Therefore, many additional instruments not identified in the SIRC reviews will exist. Second, the SIRC reviews are limited to a single healthcare setting (mental health). Third, the mapping of instruments to context constructs was based on each instrument's title/label. This can result in inaccurate mapping. For example, the Patient Safety Climate instrument [42] was mapped to climate; however, the instrument items measure cultural norms; thus, this instrument measures culture, not climate. A more appropriate approach is to map instruments based on a review of their included items. Finally, in the SIRC project, the reviewers assessed psychometric properties using the Psychometric and Pragmatic Evaluation (PAPERS) tool [43]. While this tool assesses nine psychometric criteria, it omits the most important validity criterion of content validity: the degree to which instrument items reflect the construct(s) to be measured [44]. Content validity is fundamental to assessing the validity of an instrument because lack of it can decrease internal consistency, structural validity, and interpretability of scores obtained from the instrument [45].

In this paper, we report a scoping review of instruments measuring meso- and macro-level attributes of context important to implementation by healthcare professionals, as captured in the ICON framework. We undertook a scoping review to map the breadth and depth of relevant instruments and to highlight gaps in the available instruments [46]. Our purpose was to identify psychometrically sound and pragmatic instruments for measuring macro- or meso-level ICON attributes. The review objectives were to: (1) systematically search and critically review published and unpublished literature for quantitative instruments of meso-level (organizational)

and macro-level (external to the organization, including between organizations) context attributes used in healthcare implementation efforts and (2) assess the development, psychometric (reliability, validity), and pragmatic (acceptability, easiness) properties of the identified instruments across healthcare sectors. In this paper, we report the properties of the included instruments measuring one of the macro- or meso-level ICON attributes.

Methods

Design

In the reporting of this scoping review, we adhered to the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [46] and the PRISMA-COSMIN for Outcome Measurement Instruments guideline—a consensus-based reporting standard for reviews of health measurement tools [47] (Additional file 1). The protocol is available in the Open Science Framework [48]. We conducted the review in two stages: (1) identification of instruments that align with ICON attributes and mapping them to the 20 meso- and macro-level attributes in ICON; (2) assessment of the psychometric and pragmatic properties of the identified instruments.

Stage 1: identification of instruments and mapping to the ICON's context attributes

An experienced information specialist (TR) developed and applied a search strategy to five electronic databases (OVID MEDLINE, OVID PsycInfo, Ebsco CINAHL, Ovid Health and Psychosocial Instruments, and Wiley Cochrane Library) on October 6, 2020, which was PRESS reviewed [49]. A filter for reviews [50, 51] was applied because of the large number of citations captured by the search strategy ($N=17,876$). Grey literature was not specifically searched; however, dissertations retrieved during the searches were reviewed. The search strategy and peer-review assessment are in Additional file 2.

We used EndNote [52] to deduplicate records before uploading into DistillerSR [53] for screening and data extraction. We included studies that used an instrument to measure one or more ICON attribute(s). We excluded qualitative studies, study protocols, conference abstracts, editorials, opinion papers, case studies, and studies not published in English (Table 1). Two team members (DS, LA, MDV, or WJS) independently screened the title and abstract of each record. Test screening of 10% of the full-text articles ($n=100$ articles) was conducted by two reviewers (DS, JZ, LA, MDV, or WJS) until a concordance rate of greater than 80% was achieved. The remaining records were equally divided among the reviewers for full-text screening.

Data extraction and full-text screening occurred simultaneously. Data extraction was performed by one

Table 1 Eligibility criteria

Stage 1 search	Stage 2 search
Inclusion criteria • Articles reporting studies that included an instrument measuring one or more ICON attribute(s) Exclusion criteria • Qualitative and case study designs • Study protocols, conference abstracts, editorials, opinion papers • Articles not published in English	Inclusion criteria • Articles reporting on the validity and reliability of the included instruments used in any healthcare setting and sector • Articles that described the development of instruments, regardless of the context (i.e., even if they did not originate in healthcare) Exclusion criteria • Synthesis and case study designs • Articles reporting studies that did not evaluate the instrument’s psychometric properties • Study protocols, conference abstracts, editorials, opinion papers • Articles not published in English • Articles not available in full text • Articles reporting instruments for which permission and payment was required

reviewer (DS, JZ, LA, MDV, or WJS) and verified by a second reviewer (DS or LA). We evaluated all quantitative primary studies in each included review for an instrument that potentially measured any ICON context attribute. We used a standardized and pre-tested data extraction form to extract the instrument name, the cited development article for the instrument, and the source of the instrument. For development articles in which instrument items were not reported, Google searches were performed to locate reports of the items. We excluded instruments that were not publicly available.

Two or more team members (AMH, IDG, JES, JZ, LA, LKC, MDV, and WJS) mapped each individual item of an included instrument to an ICON framework: (1) domain, (2) attribute, and (3) feature (if possible). Conflicts were resolved through consensus or consultation with senior team members (AMH, IDG, and JES). Instruments were divided as follows: (1) instruments composed of items measuring a single ICON attribute (e.g., leadership); (2) instruments composed of subscales measuring different ICON attributes; (3) instruments composed of individual items measuring different ICON attributes. Because of the very large number of instruments identified, in this review, we report only on the first category—instruments composed of items measuring a single ICON attribute.

Stage 2: search and assessment of psychometric and pragmatic properties

Search for psychometric articles of the identified instruments
From inception until August 2022, instruments were searched in PubMed using the Outcomes Measurement Properties (OMP) search filter [54]. OMP has been validated as a precise and sensitive method for searching for psychometric articles on Patient Reported Outcomes Measures (PROMs) [54] and non-PROM instruments in PubMed [55]. If the OMP search filter did not yield relevant articles and/or the instrument did not have a name,

we performed a citation search in Scopus and Web of Science using the instrument’s development article retrieved during Stage 1 and simultaneously conducted a forward search.

Eligibility criteria, article selection, data extraction

We included (1) articles that reported on the validity and reliability of the included instruments used in any healthcare setting and sector and (2) articles describing the development of instruments, regardless of the context (i.e., even if they did not originate in healthcare). We excluded synthesis studies, study protocols, conference abstracts, editorials, opinion papers, case studies, articles not published in English, articles not available in full text, or studies that did not evaluate the instrument’s psychometric properties (Table 1). Additionally, articles reporting instruments for which permission and payment were required were excluded.

We used Endnote [52] to deduplicate records before uploading them to DistillerSR [53] for screening and data extraction. Two team members (LKC, MG, VJW, or WJS) independently completed title and abstract screening, followed by full-text screening of each record. We developed a standardized data extraction form according to the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist [56, 57], which is considered best practice in the conduct of systematic reviews. Our data extraction form included: (1) instrument description, (2) study characteristics, (3) study setting, (4) sample demographics, and (5) psychometric property evidence and rating (sufficient, insufficient, indeterminate, and inconsistent) for: (a) content validity, (b) structural validity, (c) internal consistency, (d) cross-cultural validity and measurement invariance, (e) reliability, (f) measurement error, (g) hypothesis testing (discriminant, convergent, discriminative, and known groups validity), and (h) responsiveness. Data extraction

was performed independently by two team members (LKC, MG, VJW, or WJS) and in duplicate for 10% of included articles to ensure comprehensiveness of the data extraction form. After the first 10%, one reviewer extracted data from the remaining articles (LKC, MG, VJW, or WJS), which were cross-checked by a second reviewer (WJS or MG). Conflicts in screening and data extraction were resolved through consensus or consultation with a senior researcher (AMH, IDG, JES).

Assessment of measurement (psychometric) properties

In accordance with the COSMIN guidelines, a descriptive synthesis of the extracted measurement properties was performed, comprising psychometric properties and quality of evidence [56, 57]. For each instrument, we used the COSMIN criteria to evaluate development and content validity, starting with the development article and then all subsequently published articles that evaluated all criteria or any single content validity criterion (relevance, comprehensiveness, and comprehensibility) [56]. We ascribed a rating of sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?) for each content validity criterion, guided by the COSMIN manual [56, 57]. We followed this process for all other psychometric properties included in COSMIN and reported when an article utilized a variant or modified version of an instrument. This evaluation was conducted by one team member (LKC, MG, VJW, or WJS) and verified by a second reviewer (LKC, MG, VJW, or WJS). Conflicts were resolved through consensus or consultation with a senior researcher (AMH, IDG, JES).

Synthesis of the psychometric findings was organized by healthcare sectors in accordance with COSMIN recommendations, with a threshold of 85% [56] for validity evaluations and 75% for the remaining psychometric properties [57]. We then pooled psychometric properties based on healthcare sectors, regardless of the instruments' versions (i.e., some instruments had different versions).

To assess the quality of the psychometric evidence reported, we first completed the risk of bias questions outlined in the COSMIN manual for each psychometric property (from inadequate to very good) [56, 57]. We then followed the GRADE approach, modified for COSMIN [57], to summarize the quality of evidence for the psychometric properties of each instrument in a particular healthcare sector (graded from very low to high). Based on the modified GRADE approach [57], the quality of evidence was initially assumed to be high, and then lowered based on evidence of risk of bias, inconsistency, imprecision, and indirectness.

The final component of our psychometric synthesis was to categorize the recommendation level of each

instrument as A, B, or C as prescribed by COSMIN. Instruments in Category A have evidence of sufficient content validity and internal consistency and are recommended for use. Category B instruments require further psychometric testing but may be recommended for use. Category C instruments have strong evidence of insufficient measurement properties and should not be recommended for use [57].

Assessment of pragmatic properties

The pragmatic properties of the included instruments were evaluated and synthesized according to criteria outlined by Stanick and colleagues [43] on a − 1 to 4 rating scale (with − 1 being poor and 4 being excellent). The criteria used addressed acceptability (including the cost to use the instruments and use of accessible language) and easiness (including assessor burden and the length of the instrument). One team member (LKC, VJW, or WJS) evaluated the objective pragmatic criteria for the included instruments, and a second team member (LKC, VJW, or WJS) verified them. Conflicts were resolved through discussion or consultation with a senior researcher (AMH, IDG, JES).

Results

Stage 1: identification of instruments and mapping to the ICON's context attributes

Our Stage 1 search using the review search filter resulted in 1132 records, of which 499 were selected for full-text screening (Fig. 2). We included 194 reviews in our final Stage 1 sample, from which we identified 748 potential instruments. Two hundred fifty-one of these instruments could be used without permission and at no cost; of these, 180 were mapped to an ICON attribute. In total, 38 instruments were homogeneous with respect to ICON attributes, meaning they were composed of items that mapped to a single ICON attribute. All 38 instruments mapped to a meso-level ICON attribute; no instruments mapped to macro-level ICON attributes. Figure 2 shows a PRISMA diagram summarizing instrument identification and ICON mapping.

The 38 homogeneous instruments included in this review captured 9 of the 15 meso-level ICON attributes, including: (1) organizational climate ($n=3$, 7.9%); (2) physical and technological resources ($n=3$, 7.9%); (3) internal relationships ($n=6$, 15.8%); (4) organizational culture ($n=3$, 7.9%); (5) organizational change processes ($n=1$, 2.6%); (6) receptivity to change ($n=2$, 5.1%); (7) evaluation activity ($n=11$, 28.9%); (8) leadership ($n=7$, 18.4%); (9) organization of work ($n=2$, 5.3%). Figure 3 illustrates the distribution of instruments amongst the macro- and meso-level ICON framework attributes. In Table 2, we present the 38 instruments according to

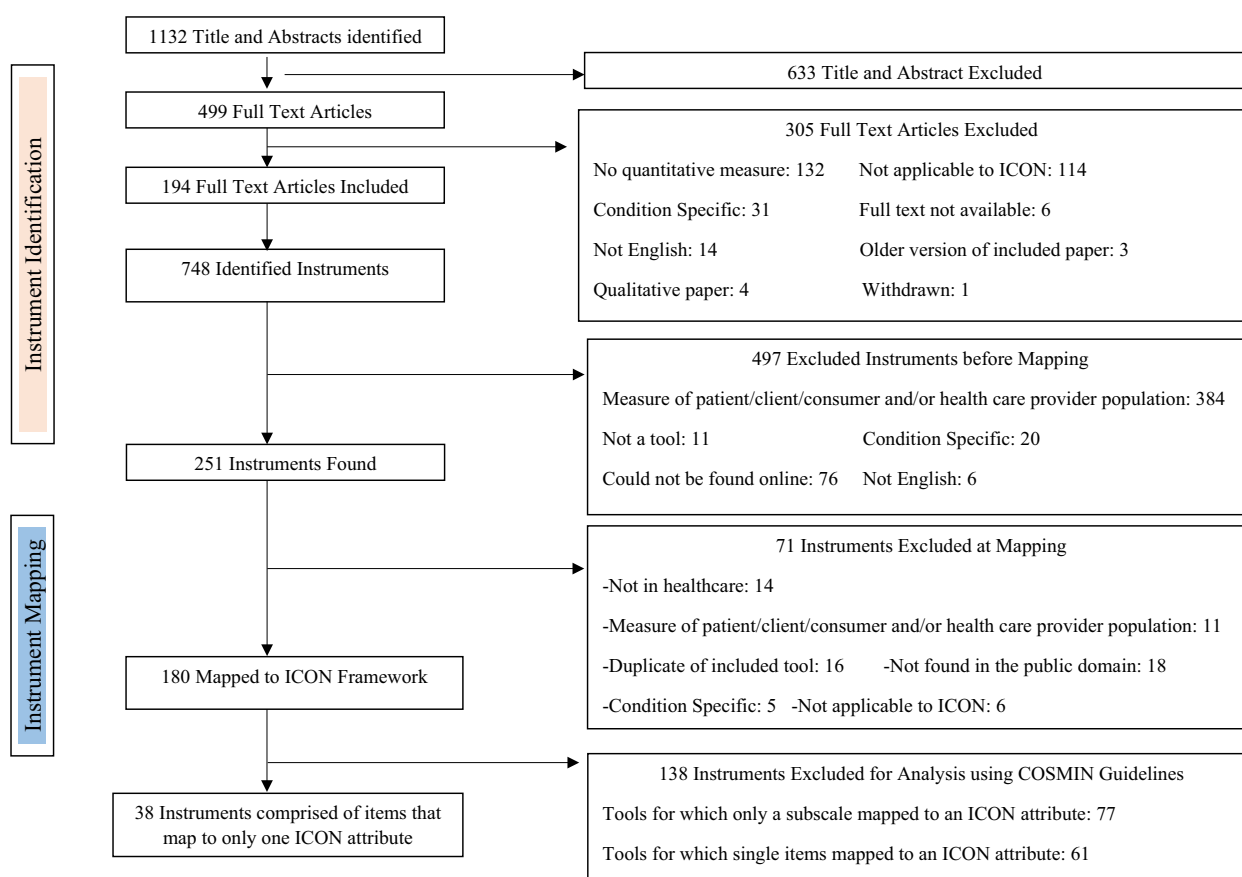


Fig. 2 PRISMA diagram identifying tools relevant to ICON

the four ICON framework meso-level domains and the nine attributes targeted in the instruments, including descriptions of instrument origins, the original and translated languages of the instruments, and the number of subscales and items for each instrument, based on their development articles. The table is organized first by ICON domain and then by each domain's respective attributes.

Stage 2: search and assessment of psychometric and pragmatic properties

Search results

Our Stage 2 searches yielded 17,450 records reporting psychometric properties across the 38 included instruments. We screened 16,377 records after removing duplicates and assessed 623 full-text articles for inclusion (Fig. 4). Included were 169 articles, with more than half ($n=102$, 60.4%) published between 2012 and 2022 (Additional file 3 lists excluded articles and the reasons). Most articles were published in Europe ($n=68$, 40.2%), followed by North America ($n=65$, 38.5%), primarily in the United States of America ($n=51$, 30.2%), then Asia

($n=28$, 16.6%), South America ($n=8$, 4.7%), Oceania ($n=7$, 4.1%), and Africa ($n=1$, 0.6%).

Assessment of instrument development

In Table 3, we present the findings on the quality of development of the 38 instruments using each instrument's development study. We located the development study for most ($n=35$, 92%) instruments, except for the following: (1) Questionnaire on Computer Systems and Decision-Making [58], (2) American Board of Internal Medicine's (ABIM) Questionnaire [59, 60], and (3) Leadership Behavior Description Questionnaire [61]. Although, we didn't locate the development studies for these three instruments, we provided references in Table 3 that describe them. In total we included 37 development articles in our assessment of 35 instruments, as the Team Climate Inventory (TCI) [62, 63] and the Interpersonal Processes of Care: IPC-29 [64, 65] had two articles each describing their development. Based on the COSMIN guidelines [56], for content validity, most instruments were rated as either inadequate ($n=16$, 42.1%) [66–81] or doubtful ($n=15$,

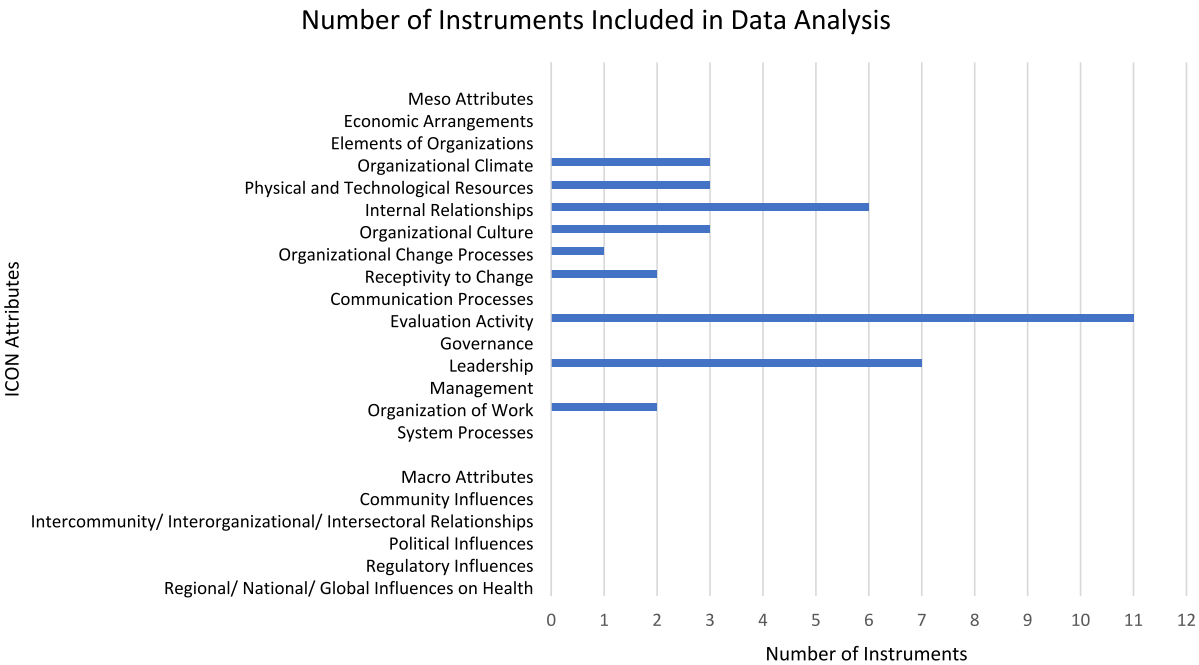


Fig. 3 Number of instruments included according to ICON attributes

39.5%) [42, 62–65, 73, 82–93]. Of the instruments with a development study, two (5.3%) were rated to have very good evidence of content validity—(1) Community Impacts of Research Oriented Partnerships (CIROP) [94] and (2) CollaboRATE Measure [95]—while two (5.3%) were rated to have adequate evidence of content validity—(1) Team Decision-Making Questionnaire [96] and (2) Care Transitions Measure (CTM)[97].

Articles reporting the development study for 35 of the included instruments clearly described the intended construct, the target population for which the instrument was developed, and the context for its use. For most instruments, the gaps in content validity were due to a lack of detailed reporting of the methods used and/or results for the items’ relevance ($n=30$, 78.9%), comprehensibility ($n=29$, 76.3%), and comprehensiveness ($n=29$, 76.3%). Furthermore, only half of the development studies ($n=19$, 50%) included a cognitive interview or pilot study with the target sample relevant to the instruments’ origin. In 16 (42.1%) of the instruments, the sample was not described in sufficient detail and/or included only topic experts and not the target population for which the instrument was developed. See Table 3 for further details of our content validity assessment of each instrument.

Measurement (psychometric) properties of the 38 instruments

In Table 4, we summarize the psychometric properties of each instrument, organized first by ICON attribute and second by the healthcare sectors in which they were evaluated. In the table, we include the number of articles reporting a psychometric property, the overall rating of each psychometric property, and the quality of evidence pertaining to the psychometric rating. Next, we provide a descriptive summary of the data, organized according to the targeted ICON attribute. In total, we summarized 169 articles for the 35 instruments for which we were able to locate psychometric studies. The number of psychometric articles varied for each of the instruments (ranging from one article to 35 articles) and across the sectors in which instruments were tested; these are noted in Table 4.

Organizational Climate. Three instruments [62, 63, 66, 82] assess *Organizational Climate* (Table 4). Organizational climate is defined as “...shared recurring patterns of behaviour, attitudes and feelings which characterize life in that organization” [25]. We mapped the three instruments into the organizational climate attribute because they measured aspects of team climate, which is a feature of organizational climate within the ICON framework [25] (see Table 2 for the feature related to each instrument). The following instruments are appropriate

Table 2 Description of instruments

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Domain: Organizational Climate and Structures									
Organization Climate									
Costa and Anderson (2011)'s [82] measure of trust in teams	1 [82]	Costa, 2011 [82]	Team Climate	Sector: Acute or Specialty Care [82] Setting: Hospital and social care institutions [82] Country: The Netherlands [82]	Employees of hospitals and social care institutions [82]	English [82]	(1) Propensity to trust (2) Perceived trustworthiness (3) Cooperative behaviours (4) Monitoring behaviours	21 [82]	7-Point Likert scale (completely disagree to completely agree) [82]
Internal Participation Scale	1 [66]	Körner, 2013 [66]	Team Climate	Sector: Rehabilitation [66] Setting: Inpatient medical rehabilitation clinics [66] Country: Germany [66]	Service providers and patient or consumer population [66]	English [66]	Unidimensional scale [66]	6 [66]	4-Point Likert scale (does not apply at all to fully applies) [66]
Team Climate Inventory (TCI)	8 [62, 63, 98–101, 103, 205]	Anderson et al. (1996, 1998) [62, 63]	Team Climate	Sector: Acute or Specialty Care [62, 63, 98–100]; Home or Community Care [205]; Primary Care [101, 103] Setting: Hospital [62, 63, 98]; Broader government and social sector [99, 100]; Home-based elderly care [205]; Primary care [101, 103] Country: Brazil [103], Canada [101], Finland [99], the Netherlands [98, 205], United Kingdom [62, 63, 100]	Service providers [63, 99–101, 103, 205], quality improvement teams [98], Management [62]	English [62, 63, 100] (Dutch [98, 205], Portuguese [103], Finish [99], French [101])	(1) Vision (2) Participatory safety (3) Support for innovation (4) Task orientation Interaction Frequency	Original: 38 [63] Short form: 14 [205]	Vision: 7-point Likert scale (1 = not at all to 7 = completely) Participative safety: 5-point response scale (1 = a very little extent to 5 = a very great extent) Task orientation: 5-point response scale (1 = strongly disagree to 5 = strongly agree) Support for innovation: 5-point response scale (1 = strongly disagree to 5 = strongly agree) [62, 63]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Physical and Technological Resources									
Questionnaire on Computer Systems and Decision-Making	0	Hatcher, 1997 [206]	Technology and Equipment & Supplies	Sector: Acute or Specialty care [206] Setting: Hospitals [206] Country: United States of America [206]	Information system personnel at hospitals [206]	English [206]	(1) Facility information (2) Information system approach (3) Information systems design (4) Impact of information systems (5) Face-to-face meetings (6) Decision-making Internet usage	41 [206]	A mixture of open-ended questions, yes or no questions, and 5-point Likert scale (not at all to quite extensively) [206]
Murphy et al. (1994)'s [68] measure of nurses' perception of a patient care information system	1	Murphy, 1994 [68]	Technology and Equipment & Supplies	Sector: Acute or Specialty care [68] Setting: Hospitals [68] Country: United States of America [68]	Service providers (Nurses) [68]	English [68]	Unidimensional scale [68]	12 [68]	5-Point Likert scale (description of scale was not reported) [68]
Perceived Health Web Site Usability Questionnaire (PHWSUQ)	2 [67, 104]	Nahm, 2006 [67]	Online Resources	Sector: Other [67, 104] Setting: Not identified [67] and inpatient and outpatient settings within a large integrated health care system [104] Country: United States of America [67, 104]	Patient or consumer population [67, 104]	English [67, 104]	(1) Satisfaction (2) Ease of reading (3) Usefulness	10 [67, 104]	7-Point Likert scale (very unsatisfied to very satisfied) [67]
Domain: Organizational Social Behaviour									
Internal Relationships Collaborative Practice Scale	2 [69, 207]	Weiss, 1985 [69]	Partnerships (or Collaborative Practices)	Sector: Acute or Specialty care [69, 207] Setting: Hospitals [69, 207] Country: Switzerland [207]; United States of America [69]	Service providers (Nurses and Physicians) [69, 207]	English [69] (German [207])	(1) Nurses (2) Physicians	19 [69, 207]	6-Point Likert scale (always to never) [69, 207]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Community Impacts of Research Oriented Partnerships (CIROP)	1 [94]	King 2009 [94]	Partnerships (or Collaborative Practices)	Sector: Home or community [94] Setting: Community-based health research [94] Country: Canada [94]	Researchers and community partners [94]	English [94]	(1) Personal knowledge development (2) Personal research skill development (3) Organizational/group access to and use of information (4) Community and organizational development	33 [94]	7-Point Likert scale (not at all to a very great extent) [94]
Multiple-group measurement scale for inter-professional collaboration	2 [70, 208]	Kenaszchuk, 2010 [70]	Partnerships (or Collaborative Practices)	Sector: Acute or Specialty care [70, 208] Setting: Hospitals [70, 208] Country: Canada [209], Italy [208]	Service providers [70, 208]	English [70] (German [208], Italian [208])	(1) Communication (2) Accommodation (3) Isolation	13 [70]	4-Point Likert scale (strongly disagree to strongly agree) [70]
Shared Decision-Making Questionnaire-Doc (SDM-Q-Doc)	7 [71, 105, 127, 128, 137, 210, 211]	Scholl, 2012 [71]	Partnerships (or Collaborative Practices)	Sector: Acute or Specialty care [105, 127, 128, 210]; Primary care [71, 137, 211] Setting: Hospital [105, 210]; Oncology clinical settings [127]; Specialty outpatient clinics (e.g., psychiatry, oncology etc.) [128]; Primary care clinics [71, 137, 211] Country: Iran [127], Japan [105, 211], Germany [71, 137], Spain [210], the Netherlands [128]	Service providers (specifically physicians) [71, 105, 127, 128, 137, 210, 211]	German [71, 137] (English [71, 137], Japanese [105, 211], Spanish [210], Persian [127], Dutch [128])	Unidimensional scale [71]	9 [71]	6-Point Likert scale (completely disagree to completely agree) [71]
Team Decision-Making Questionnaire	1 [96]	Batorowicz, 2008 [96]	Partnerships (or Collaborative Practices)	Sector: Acute or Specialty [96] Setting: Augmentative and Alternative Communication (AAC) Centres [96] Country: Canada [96]	Service providers [96]	English [96]	Unidimensional scale [96]	19 [96]	7-Point Likert scale (not at all to a vast extent) [96]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Trust in Health Promotion Partnerships Scale	1 [83]	Jones, 2011 [83]	Partnerships (or Collaborative Practices)	Sector: Home or community care [83] Setting: Health promotion partnerships (hospitals, community health services, voluntary groups, etc.) [83] Country: Ireland [83]	Service providers [83]	English [83]	Unidimensional scale [83]	14 [83]	5-Point Likert scale (never to always) [83]
Organizational Culture									
On-line Patient Safety Climate Survey	1 [42]	Connelly, 2005 [42]	Cultural Norms	Sector: Acute or Specialty [42] Setting: Military hospitals [42] Country: United States of America [42]	Service providers [42]	English [42]	(1) Willingness to report errors (2) Problem-solving processes (3) Attitude of leadership	19 [42]	4-Point Likert scale (strongly disagree to strongly agree) [42]
Stanford University Patient Safety Climate in Healthcare Organizations questionnaire	1 [84]	Singer, 2007 [84]	Cultural Norms	Sector: Acute or Specialty [84] Setting: Hospitals [84] Country: United States of America [84]	Service providers and management [84]	English [84]	(1) Senior Managers' engagement (2) Organizational resources for safety (3) Overall emphasis on safety (4) Unit safety norms (5) Unit recognition and support for safety efforts (6) Fear of shame (7) Provision of safe care (8) Learning (9) Fear of blame	38 [84]	5-Point Likert scale (never to always) [84]
Survey on Patient Safety in Ambulatory Care Organizations	1 [85]	Schutz, 2007 [85]	Cultural Norms	Sector: Acute or Specialty [85] Setting: Ambulatory clinics (paediatric, surgery, women's health, etc.) [85] Country: United States of America [85]	Service providers [85]	English [85]	(1) Teamwork (2) Leadership (3) Communication (4) Tendency to Report Errors (5) Global Question	21 [85]	5-Point Likert scale (strongly disagree to strongly agree) [85]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Domain: Organizational Response to Change									
Organizational Change Processes									
Howell and colleagues [72] measure of champion behaviour	1 [72]	Howell, 2005 [72]	Champions/Opinion Leaders	Sector: Non-healthcare [72] Settings: Business settings [72] Country: United States of America [72]	Champions of manufacturing firms (electronics, automotive, and aerospace industries) [72]	English [72]	(1) Risk-taking propensity (2) Taking chances, innovativeness (3) Develops novel solutions to problems and achievement orientation	15 [72]	5-Point Likert scale (not at all to always) [72]
Receptivity to Change									
Organizational Readiness for Implementing Change (ORIC) scale	6 [86, 107–111]	Shea, 2014 [86]	Readiness for Change	Sector: Acute care or specialty [86, 107, 108]; Home or community [110]; Primary care [109, 111] Settings: Cancer services [107], Hospitals [86], Healthcare education (university) [86], Home care services [110] Obstetrics [108], Primary care [109, 111] Country: Brazil [109], Canada [110], Denmark [108], Germany [107], South Africa [111], United States of America [86]	Service providers [86, 107–111]	English [86, 110, 111] (Danish [108], French [110], German [107], Portuguese [109])	(1) Effectiveness (2) Commitment	12 [86]	5-Point Likert scale (disagree to agree) [86]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Resistance to Change Scale	2 [73, 112]	Oreg, 2003 [73]	Compatibility and Readiness for Change	Sector: Other [73, 112] Setting: Universities [73, 112] Country: 17 countries [73] (Australia, China, Croatia, the Czech Republic, Germany, Greece, Israel, Japan, Lithuania, Mexico, the Netherlands, Norway, Slovakia, Spain, Turkey, the United Kingdom, and the United States of America), United States of America [112]	Service providers [73, 112]	English [112] (Chinese, Croatian, Czech, German, Greek, Hebrew, Japanese, Lithuanian, Spanish, Dutch, Norwegian, Slovakian, Spanish, Turkish) [73]	(1) Routine seeking (2) Emotional reaction (3) Short-term focus (4) Cognitive rigidity	16 [73]	6-Point Likert scale (strongly disagree to strongly agree) [73]
Domain: Organizational Processes									
Evaluation Activity									
American Board of Internal Medicine's (ABIM) questionnaire	0	Webster, 1989 [59, 212]	Patient/Client/Consumer Feedback to Staff	Sector: Acute care or specialty [212] Setting: Hospitals [212] Country: United States of America [212]	Patient/Client/Consumer Population [212]	English [212]	Unidimensional	Complete: 23 [212] Short: 10 [212]	5-Point Likert scale (poor to excellent) [212]
Care Transitions Measure (CTM)	12 [97, 172–180, 182, 213]	CTM-15: Coleman, 2002 [97] CTM-3: Parry, 2008 [175]	Patient/Client/Consumer Feedback to Staff	Sector: Acute care or specialty [172–180, 213]; Primary care [97]; Rehabilitation [182] Setting: Hospitals [172–177, 179, 180, 213], oncology centre [178]; Primary care clinics [97]; Rehabilitation units [182] Country: Brazil [172]; Canada [182]; China [177]; Israel [178]; Japan [213]; Korea [180]; Singapore [176]; Sweden [174]; United States of America [97, 173, 175, 179]	Patient/Client/Consumer Population [97, 175]	English [97, 173, 175, 179, 182] (Arabic [178]; Chinese [176, 177]; Hebrew [178]; Japanese [213]; Korean [180]; Swedish [174]; Portuguese [172])	(1) Critical understanding [6] (2) Preferences important [3] (3) Management preparation [4] (4) Care plan [2]	CTM-15: 15 [97] CTM-3: 3 [175]	4-Point Likert scale (strongly disagree to strongly agree) [97, 175]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
CollaboRATE measure	11 [95, 125, 214–222]	Elwyn, 2013 [95]	Patient/Client/Consumer Feed-back to Staff	Sector: Acute care or specialty [95, 125, 214–219]; Home or community [220]; Primary care [221]; Unidentified [222] Setting: Ambulatory primary care clinics [221]; Cancer research [217]; CPAP clinics [215]; Community mental health services [220]; Hospitals [95, 214, 216, 219]; Orthopaedic specialist clinics [218]; Paediatric ambulatory clinics [125] Country: Argentina [214]; Canary Islands [220]; France [217]; Sweden [215]; the Netherlands [216]; United States of America [95, 125, 218, 219, 221, 222]	Patient/Client/Consumer Population [95]	English [95, 218, 221, 222] (Dutch [216]; French [217]; Spanish [125, 214, 219, 220]; Swedish [215])	Unidimensional scale [95]	3	9-Point Likert scale (no effort was made to every effort was made) [95]
Doctors' Interpersonal Skills Questionnaire (D/ISQ)	2 [74, 223]	Greco, 1999 [74]	Patient/Client/Consumer Feed-back to Staff	Sector: Primary care [74, 223] Setting: Primary care clinics [74, 223] Country: Australia [74]; United Kingdom [223]	Patient/Client/Consumer Population [74]	English [74, 223]	Unidimensional scale [74]	12	6-Point Likert scale (varied with the item stem; e.g., not at all satisfied to extremely satisfied) [74]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Interpersonal processes of care: IPC	9 [64, 65, 113–119]	Stewart, 2007 [65]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or Specialty care [65, 114, 115]; Primary care [116–119]; Home or community [113]; Public Health [64] Setting: Adult Medical Clinic (public sector facility) [64]; Diabetes specialty clinics [114]; Hospital [65]; Psychiatric clinics [115]; Primary care clinics [116–119]; Senior centres, community health clinics, unemployment agencies, and colleges [113] Country: Canada [116–119]; Iran [114]; Japan [115]; United States of America [64, 65, 113]	Patient/Client/Consumer Population [65]	English [64, 65, 113, 116–119] (French [117, 118]; Japanese [115]; Spanish [113]; Persian [114])	Hurried communication 1) Lack of clarity (2) Hurried, distracted (3) Elicited concerns, responded (4) Explained results (5) Explained medications Decision-Making (6) Asked patient (7) Decided together Interpersonal Style (8) Emotional support, compassion (9) Respectfulness (10) Assumed socioeconomic status (11) Discrimination due to race/ethnicity (12) Disrespectful office staff	IPC I = 41 [64] IPC II = 29 [65]	5-Point Likert scale (never to always) [64, 65]
Patient Participation Emergency Department questionnaire (PPED)	1 [87]	Frank 2011 [87]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or Specialty care [87] Setting: Hospital emergency departments [87] Country: Sweden [87]	Patient/Client/Consumer Population [87]	Swedish [87]	(1) Fight for participation (2) Requirement for participation (3) Mutual participation (4) Participating in getting basic needs satisfied	17 [87]	4-Point Likert scale (strongly disagree to strongly agree) [87]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Risser Patient Satisfaction Scale	6 [88, 224–228]	Risser, 1975 [88]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or Specialty care [224–228]; Primary care [88] Setting: Hospitals [225, 227, 228]; Oncology units [224]; Orthopaedic units [226]; Primary care clinic [88] Country: Cyprus [224]; Greece [225]; United States of America [88, 226–228]	Patient/Client/Consumer Population [88]	English [88, 226–228] (Greek [224, 225])	(1) Technical-Professional (2) Interpersonal-educational (3) Interpersonal-trusting	25 [88]	5-Point Likert scale (strongly disagree to strongly agree) [88]
Role-Based Performance Scale	1 [89]	Welbourne, 1998 [89]	Patient/Client/Consumer Feedback to Staff	Sector: Non-healthcare [89] Setting: Various industries, including traditional manufacturing, service, and high technology [89] Country: United States of America [89]	Patient/Client/Consumer Population [89]	English [89]	(1) Innovator (2) Career Role-Based Performance Scale(3) Team (4) Organization (5) Job	20 [89]	5-Point Likert scale (needs much improvement to excellent) [89]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Shared Decision-Making Questionnaire (SDM-Q-9)	21 [90, 120–139, 229]	Simon, 2006 [90]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or specialty care [90, 120–134, 139]; Primary care [90, 135–139] Setting: Breast cancer clinics [131, 132]; Community outpatient behavioural health [121]; Hospitals [90, 122, 124, 128, 129]; Multiple Sclerosis clinics [123, 126]; Maternity care [130]; Oncology units [127]; Outpatient diabetic clinic [120]; Paediatric clinics [125]; Primary care [90, 135–139]; Psychiatric clinics [133]; Specialty clinics [134, 139] Country: Ecuador [120]; Germany [90, 126, 129, 130, 137, 138]; Hungary [139]; Iran [127]; Israel [133]; Japan [136]; Malaysia [131]; the Netherlands [128]; Romania [122]; Spain [123, 124, 135]; Taiwan [132]; United States of America [121, 125]; United Arab Emirates [134]	Patient/Client/Consumer Population [90]	German [90, 126, 130, 137, 138] (Arabic [134]; Chinese [131, 132]; Dutch [128]; English [121, 125, 131]; Japanese [136]; Hebrew [133]; Hungarian [139]; Malay [131]; Persian [127]; Romanian [122]; Spanish [120, 121, 123–125, 135])	Unidimensional scale [90]	9 [90]	6-Point Likert scale (completely disagree to completely agree) [90]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
The Patient Assessment of Chronic Illness Care Survey (PACIC)	35 [91, 140–171, 230, 231]	Glasgow, 2005 [91]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or specialty [140–150, 163, 165–168, 231]; Primary care [91, 151–160, 164, 165, 169–171, 230]; Long-term care [165]; Home or community [165]; Public health [165]; Rehabilitation [161]; Unidentified [162] Setting: Cardiovascular disease care [142]; Cardiac rehabilitation sites [161]; Diabetes clinics [141, 143, 144, 148, 149, 166, 168, 231]; Heart transplant centres [163]; Hospitals [140, 145–147, 150, 167]; Primary care clinics [91, 152–160, 164, 169–171, 230]; Varied settings (general practitioner office, hospices, public health, hospitals etc.) [165] Country: Australia [141, 158, 163]; Belgium [163]; Brazil [163]; Canada [161–163]; Denmark [148]; Finland [149]; France [147, 163]; Germany [154, 163–165, 170, 171]; Greece [167]; India [146]; Italy [163]; Malaysia [151, 152]; New Zealand [230]; the Netherlands [142, 150, 160, 169, 230]; Saudi Arabia [166]; Spain [163]; Slovenia [159]; South Korea [145]; Switzerland [144, 163]; Thailand [168]; United Kingdom [155, 157, 163]; United States of America [91, 140, 143, 153, 156, 163, 231]	Patient/Client/Consumer Population [91]	English [91, 141, 143, 153–155, 161, 230, 231] (Arabic [166]; Dutch [142, 163, 169]; French [144, 163]; German [163, 164]; Italian [163]; Malay [151, 152]; Portuguese [163]; Spanish [140, 163])	(1) Patient activation: (2) Delivery system design/decision support (3) Goal setting/tailoring (4) Problem solving/contextual (5) Follow-up/coordination	PACIC = 20 [91] PACIC short = 11 [164] PACIC5A = 26 [171] PACIC with culturally sensitive items = 27 [230]	5-Point Likert scale (almost never to almost always) [91]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Urgent Care System Questionnaire (UCSQ)	1 [92]	O' Cathain, 2011 [92]	Patient/Client/Consumer Feedback to Staff	Sector: Acute or specialty [92] Setting: Hospitals [92] Country: England [92]	Patient/Client/Consumer Population [92]	English [92]	(1) Progress (2) Entry (3) Patient convenience	18 [92]	5-Point Likert scale (strongly disagree to strongly agree) [92]
Leadership									
Ehrhart 2004 [75] Servant Leadership Scale	2 [75, 232]	Ehrhart, 2004 [75]	Leadership style	Sector: Home or Community [232]; Non-healthcare Setting: Social Welfare Organizations [232]; Grocery department stores [75] Country: Korea [232]; United States of America [75]	Service providers [75]	English [75, 232]	Unidimensional scale [75]	14 [75]	5-Point Likert scale (to a very small extent to a great extent) [75]
Global Transformational Leadership Scale	2 [76, 233]	Carless, 2000 [76]	Leadership style	Sector: Acute or Specialty care [233]; Non-healthcare [76] Setting: Hospitals [233]; Retail bank [76] Country: Pakistan [233]; Australia [76]	Service providers [76]	English [76, 233]	Unidimensional scale [76]	7 [76]	5-Point Likert scale (rarely or never to very frequently, if not always) [76]
Implementation Leadership Scale	9 [77, 183–188, 234, 235]	Aarons, 2014 [77]	Leadership style	Sector: Acute or Specialty care [77, 183–188]; Home or community [234, 235] Setting: Alcohol and other drug treatment [183]; Child welfare community-based organization [234]; Hospitals [184–187]; Health services in schools [235]; Mental health programme [188]; Outpatient mental health [77] Country: China [184, 185]; Greece [186]; United States of America [77, 183, 187, 188, 234, 235]	Service providers [77]	English [77, 183, 187, 188, 234, 235] (Chinese [184, 185]; Greek [186])	(1) Proactive leadership (2) Knowledgeable leadership (3) Supportive leadership (4) Perseverant leadership	12 [77]	5-Point Likert scale (not at all to very great extent) [77]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Leadership Behavior Description Questionnaire	0	Stogdill, 1963 [236]	Leadership style	Sector: Non-health care [236] Setting: Diverse setting (commissioned and non-commissioned police officers, executives in aircraft engineering staff, presidents of labour unions, presidents of colleges and universities, etc.) [236] Country: United States of America [236]	Service providers [236]	English [236]	(1) Initiating Structure (2) Consideration	15 [236]	5-Point Likert scale (never to always) [236]
Servant Leadership Questionnaire	1 [78]	Barbuto & Wheeler, 2006 [78]	Leadership style	Sector: Non-health care [78] Setting: Community leaders [78] Country: United States of America [78]	Service providers [78]	English [78]	(1) Altruistic calling (2) Emotional healing (3) Wisdom (4) Persuasive mapping (5) Organizational stewardship	23 [78]	4-Point Likert scale (strongly disagree to strongly agree) [78]
Servant Leadership Survey	3 [79, 189, 190]	van Dierendonck, 2010 [79]	Leadership style	Sector: Non-health care [189]; Not reported [79, 190] Setting: Private companies, non-profit organizations, or government agencies/public organizations [189]; Not reported [79, 190] Country: Japan [189]; Argentina [190]; Spain [190]; Mexico [190]; the Netherlands [79]; United Kingdom [79]	Service providers [79]	Dutch [79] English [79] (Japanese [189]; Spanish [190])	(1) Empowerment (2) Accountability (3) Standing back (4) Humility (5) Authenticity (6) Courage (7) Interpersonal acceptance (8) Stewardship	30 [79]	7-Point Likert scale (strongly disagree to strongly agree) [79]
The Mentorship Effectiveness Scale	1 [80]	Berk, 2005 [80]	Mentors	Sector: Healthcare education [80] Setting: Educational institutions [80] Country: United States of America [80]	Service providers (nursing faculty) [80]	English [80]	(1) Description of Relationship (2) Outcome Measures	12 [80]	Open-ended questions and select all that apply [80]

Table 2 (continued)

Instrument	No. of articles	Reference of development article (s)	ICON feature	Sector; setting; country	End user of instrument	Original language (translated languages)	Subscales	No. of items	Response options
Organization of Work									
A107 NASA Task Load Index	6 [81, 191, 192, 237–239]	Hart and Staveland, 1988 [81]	Workload	Sector: Acute care or specialty [191, 192, 238]; Healthcare Education [237, 239] and Non-healthcare (Industry) [81] Setting: Intensive care departments [192, 238]; Diabetes management [191]; Healthcare education—University surgical training setting [237, 239]; National Aeronautics and Space Administration (NASA) [81] Country: United States of America [81, 191, 192, 237, 238]; Canada [239]	Service providers [191, 192, 238]	English [81, 191, 192, 237–239]	(1) Mental Demand (2) Physical Demand (3) Temporal Demand (4) Performance (5) Effort (6) Frustration	6 [81]	7-Point scale (very low to very high with 21 gradations) [81]
RAFAELA System: (1) Professional Assessment of Nursing Care Intensity Level (PAONCIL) and (2) Oulu Patient Classification (OPC)	10 [93, 193–198, 240–242]	Fagerstrom & Rainio, 1998 [93]	Workload	Sector: Acute or specialty [93, 193–197, 240–242]; Home or community [198] Setting: Hospitals [93, 193, 195, 240–242]; Organ transplantation [197]; Mixed specialty units from 8 hospitals [194]; Gynaecology units [93]; Home healthcare units [198] Country: Finland [93, 193–196, 240, 241]; Norway [197, 198]; Amsterdam [242]	Service providers [93]	Finnish [93, 193–196, 240, 241] (Norwegian [197, 198], Dutch [242])	PAONCIL: Single item scale [93] OPC: (1) planning and coordination of nursing care (2) breathing, blood circulation and symptoms of disease (3) nutrition and medication (4) personal hygiene (5) activity, sleep and rest (6) teaching, guidance in care and follow-up care, emotional support [194]	PAONCIL = 2 [93] OPC = 6 item [194]	PAONCIL: Segmented line scale (very low to very high) [93] OPC: Nursing intensity classified as A (1 point), B (2 points), C (3 points), D (4 points) increasing points means increasing intensity [194]

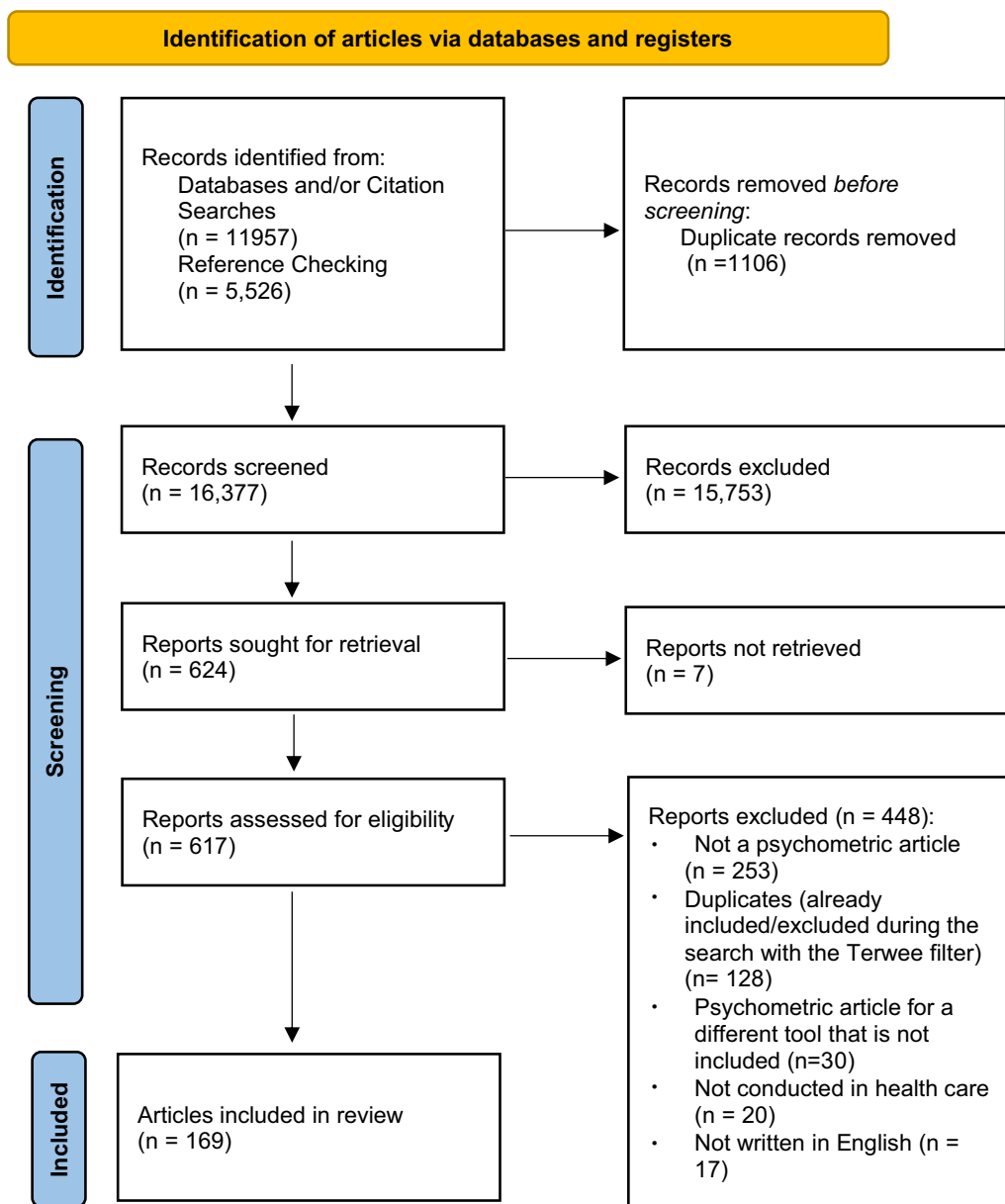


Fig. 4 PRISMA diagram

to use in the respective healthcare sectors: (1) Internal Participation Scale [66] (rehabilitation); (2) Team Climate Inventory (TCI) [62, 63, 98] (acute care/specialty care and primary care); (3) Costa and Anderson's (2011) [82] Measure of Trust in Teams (acute care/specialty care). While additional studies evaluated the psychometric properties of the TCI in home/community/public health settings [99, 100], content validity was not appraised in these sectors; therefore, the TCI cannot be recommended for use in these sectors at this time. Furthermore, several variants of the TCI were evaluated; three of

eight (37.5%) studies evaluated the short form in primary care [101], home or community care [102], and public health [99] sectors. Two of these studies confirmed, using confirmatory factor analysis, the same four-factor structure (participative safety, support for innovation, vision, and task orientation) composed of 14-items [99, 102]. Beaulieu and colleagues found a similar factor structure but only conducted an exploratory factor analysis in testing a 19-item short version of the TCI [101]. The remaining five articles reported evaluations of a long version (38

Table 3 (continued)

Instrument	Instrument design		Cognitive interview (CI) study or pilot test									
	General design requirements		General design requirements									
	Clear construct	Clear origin of construct	Clear target population for which the instrument was developed	Clear context of use	Instrument developed in sample representing the target population	Concept elicitation (relevance and comprehensiveness)	Total instrument design	CI study performed in sample representing the target population	Comprehensibility	Comprehensiveness	Total CI study	Total instrument development
Community Impacts of Research-Oriented Partnerships (CIROP) [94]	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good
Multiple-group measurement scale for interprofessional collaboration [70]	Very good	Adequate	Very good	Very good	Adequate	Doubtful	Doubtful	Inadequate	Inadequate	Doubtful	Inadequate	Inadequate
Shared Decision-Making Questionnaire-Doc (SDM-Q-Doc) [71]	Very good	Very good	Adequate	Very good	Adequate	Doubtful	Doubtful	Inadequate	Inadequate	Doubtful	Inadequate	Inadequate
Team Decision-Making Questionnaire [96]	Very good	Very good	Very good	Very good	Very good	Adequate	Adequate	Very good	Adequate	Adequate	Adequate	Adequate
Trust in Health Promotion Partnerships Scale [83]	Very good	Very good	Very good	Very good	Very good	Adequate	Adequate	Very good	Doubtful	Very good	Doubtful	Doubtful
Organizational Culture												
On-line Patient Safety Climate Survey [42]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Very good	Doubtful	Doubtful	Doubtful	Doubtful
Stanford University Patient Safety Climate in Healthcare Organizations questionnaire [84]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Doubtful	Doubtful	Doubtful	Doubtful	Doubtful
Survey on Patient Safety in Ambulatory Care Organizations [85]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Very good	Doubtful	Doubtful	Doubtful	Doubtful

Table 3 (continued)

Instrument	Instrument design		Cognitive interview (CI) study or pilot test									
	General design requirements					General design requirements						
	Clear construct	Clear origin of construct	Clear target population for which the instrument was developed	Clear context of use	Instrument developed in sample representing the target population	Concept elicitation (relevance and comprehensiveness)	Total instrument design	CI study performed in sample representing the target population	Comprehensibility in sample	Comprehensiveness	Total CI study	Total instrument development
Domain: Organizational Response to Change												
Organizational Change Processes												
Howell and colleagues [72] measure of champion behaviour	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate
Receptivity to Change												
Organizational Readiness for Implementing Change (ORIC) Scale [86]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Adequate	Doubtful	Doubtful	Doubtful	Doubtful
Resistance to Change Scale [73]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate
Domain: Organizational Processes												
Evaluation Activity												
American Board of Internal Medicine's (ABIM) questionnaire [59, 60]												
Care Transitions Measure (CTM) [97]	Very good	Very good	Very good	Very good	Adequate	Adequate	Adequate	Very good	Adequate	Very good	Adequate	Adequate
Colloborate measure [95]	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good	Very good
Doctors' Interpersonal Skills Questionnaire (DISQ) [74]	Very good	Doubtful	Very good	Very good	Adequate	Doubtful	Doubtful	Inadequate	Doubtful	Doubtful	Doubtful	Inadequate
Interpersonal processes of care: IPC-29 [64, 65]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Very good	Very good	Doubtful	Doubtful	Doubtful

Table 3 (continued)

Instrument	Instrument design				Cognitive interview (CI) study or pilot test							
	General design requirements				General design requirements							
	Clear construct	Clear origin of construct	Clear target population for which the instrument was developed	Clear context of use	Instrument developed in sample representing the target population	Concept elicitation (relevance and comprehensiveness)	Total instrument design	CI study performed in sample representing the target population	Comprehensibility	Comprehensiveness	Total CI study	Total instrument development
Servant Leadership Questionnaire [78]	Very good	Very good	Very good	Very good	Adequate	Doubtful	Doubtful	Inadequate	Doubtful	Doubtful	Inad- equate	Inadequate
Servant leadership Survey [79]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Inadequate	Doubtful	Doubtful	Inad- equate	Inadequate
The Mentorship Effectiveness Scale [80]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Inadequate	Doubtful	Doubtful	Inad- equate	Inadequate
Organization of Work												
NASA Task Load Index [81]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Inadequate	Doubtful	Doubtful	Inad- equate	Inadequate
RAFAELA System [93]	Very good	Very good	Very good	Very good	Very good	Doubtful	Doubtful	Very good	Doubtful	Doubtful	Doubt- ful	Doubtful

Table 4 Measurement property rating and adapted GRADE according to sector organized by attributes

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Domain: Organizational Climate and Structures											
Organizational Climate											
Costa and Anderson (2011)'s [82]	Acute care or specialty (n = 1) [82]	+ (n = 1); Low [82]		+ (n = 1); High [82]			+ (n = 1); High [82]			+ (n = 1); High [82]	A
measure of trust in teams											
Internal Participation Scale [66]	Rehabilitation (n = 1) [66]	+ (n = 1); Low [66]		+ (n = 1); High [66]			+ (n = 1); High [66]				A
Team Climate Inventory (TCI) (n = 3) [62, 63, 98]	Acute care or specialty (n = 3) [62, 63, 98]	+ (n = 1); Low [62]		+ (n = 2); High [63, 98]			+ (n = 3); High [62, 63, 98]		+ (n = 2); Low [63, 98]	+ (n = 1); Low [63]	A
	Primary Care (n = 2) [101, 103]	+ (n = 1); Low [103]		? (n = 2); Moderate [101, 103]			+ (n = 2); High [101, 103]		+ (n = 1); Low [103]		A
	Home or community (n = 1) [205]			+ (n = 1); High [205]			+ (n = 1); High [205]			-(n = 1); Low [205]	B
	Public Health (n = 2) [99, 100]			+ (n = 2); High [99, 100]			+ (n = 2); High [99, 100]		- (n = 1); Low [99]		B
Physical and Technological Resources											
Questionnaire on Computer Systems and Decision-Making [58]	No psychometric paper										
A207 Murphy et al. (1994)'s measure of nurses' perception of a patient of care information system	Acute or Specialty (n = 1) [68]	+ (n = 1); Low [68]		? (n = 1); Low [68]			? (n = 1) Low [68]		- (n = 1); Low [68]		B
Perceived Health Web Site Usability Questionnaire (PHWSUQ)	Public Health (n = 2) [67, 104]	? (n = 1); Low [67]		? (n = 1); Low [104]			? (n = 2); Moderate [67, 104]		+ (n = 1); Low [104]		B

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Domain: Organizational Social Behaviour											
Internal Relationships											
Collaborative Practice Scale	Acute or Speciality (n=2) [69, 207]			? (n=2); Moderate [69, 207]	? (n=2); High [69, 207]			+ (n=2); Moderate [69, 207]		± (n=1); Low [69]	B
Community Impacts of Research Oriented Partnerships (CIROP)	Home or community (n=1) [94]		+ (n=1); High [94]	? (n=1); Moderate [94]	? (n=1); High [94]			± (n=1); Low [94]	+ (n=1); Low [94]	- (n=1); Low [94]	B
Multiple-Group Measurement Scale for Inter-professional Collaboration	Acute or Speciality (n=1) [70]		? (n=1); Very low [70]	+ (n=1); High [70]	+ (n=1); High [70]			+ (n=1); High [70]		+ (n=1); Low [70]	B
Shared Decision-Making Questionnaire- Doc (SDM-Q- Doc)	Public Health (n=1) [208]		? (n=1); Low [208]	+ (n=1); High [208]	+ (n=1); High [208]						B
	Acute or Speciality (n=3) [127, 128, 210]		? (n=3); Low [127, 128, 210]	+ (n=3); High [127, 128, 210]	+ (n=3); High [127, 128, 210]			+ (n=3); Moderate [127, 128, 210]	+ (n=1); High [210]		B
Primary Care	(n=2) [137, 211]		? (n=1); Low [137, 211]	+ (n=1); High [211]	+ (n=1); High [211]			+ (n=1)		- (n=1); Low [137]	B
								+ (n=1)			
Long-Term Care	(n=2) [71, 105]		+ (n=2); Very Low [71, 105]	+ (n=2); Moderate [71, 105]	+ (n=2); Moderate [71, 105]					- (n=1); Low [71]	A
Team Decision-Making Questionnaire	Acute or speciality (n=1) [96]		+ (n=1); High [96]	? (n=1); High [96]	? (n=1); High [96]					± (n=1); Low [96]	B
Trust in Health Promotion Partnerships Scale	Public Health (n=1) [83]		+ (n=1); Low [83]	? (n=1); Moderate [83]	? (n=1); Low [83]						B
Organizational Culture											
On-line Patient Safety Climate Survey	Acute or speciality (n=1) [42]		+ (n=1); Low [42]	? (n=1); Very low [42]	? (n=1); Very low [42]						B

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Stanford University Patient Safety Climate in Healthcare Organizations questionnaire	Acute or speciality (n = 1) [84]	+(n = 1); Low [84]	?(n = 1); Moderate [84]	?(n = 1); Low [84]							B
Survey on Patient Safety in Ambulatory Care Organizations	Acute or speciality (n = 1) [85]	+(n = 1); Low [85]									B
Domain: Organizational Response to Change											
Organizational Change Processes											
Howell and colleagues [72] measure of champion behaviour	Other- Business settings (n = 1) [72]	+(n = 1); High [72]	+(n = 1); High [72]	+(n = 1); High [72]	+(n = 1); High [72]			+(n = 1); High [72]			A
Receptivity to Change											
Organizational Readiness for Implementing Change (ORIC) scale	Acute or Speciality (n = 2) [107, 108]	? (n = 2); Moderate [107, 108]	High [107, 108]	+(n = 2); Moderate [107, 108]							B
	Primary Care (n = 2) [109, 111]	? (n = 1); Low [109, 111]	Moderate [109, 111]	-(n = 2); Moderate [109, 111]	+(n = 2); Moderate [109, 111]				? (n = 1); Doubtful [111]	+(n = 1); Low [111]	B
	Home or community (n = 1) [110]	+(n = 1); Moderate [110]	High [110]	-(n = 1); High [110]	+(n = 1); High [110]						B
	Acute or Specialty, Public health, and non-health care (n = 1) [86]	+(n = 1); Low [86]	High [86]	+(n = 1); High [86]	+(n = 1); High [86]					+(n = 1); High [86]	A
Resistance to Change scale	Other- Universities (n = 2) [73, 112]	+(n = 1); Low [73]	High [73, 112]	+(n = 2); High [73, 112]	+(n = 2); High [73, 112]	+(n = 1); High [112]		+(n = 2); High [73, 112]	+(n = 1); High [73]		A

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Domain: Organizational Processes											
Evaluation Activity											
American Board of Internal Medicine's (ABIM) questionnaire [59, 60]	No psychometric paper										
	Care Transitions Measure (CTM)	Acute or Speciality (n = 10) [172–180, 213]	+(n = 6); Low [172, 174, 177, 178, 180, 213]	High [172–180, 213]	? (n = 10); High [172–180, 213]	+(n = 10); Moderate [172–180, 213]	Moderate [172, 176–178, 213]	+(n = 5); Moderate [173, 175–177, 179, 180, 213]	+(n = 7); Moderate [173, 175–177, 179, 180, 213]	+(n = 3); Low [172, 177, 180]	A
	Primary Care (n = 1) [97]					+(n = 1); Low [97]	? (n = 1); Low [97]			+(n = 1); Low [97]	B
CollaborATE measure	Rehabilitation (n = 1) [182]		+(n = 1); Moderate [182]				Very low [182]	+(n = 1); Very low [182]			B
	Acute or Speciality (n = 8) [95, 125, 214–219]	+(n = 5); Moderate [95, 125, 214, 217, 219]	+(n = 2); Moderate [95, 125, 214, 217, 219]	? (n = 2); High [125, 214, 215, 217]			Moderate [95, 125, 214–216, 218]	+(n = 6); Moderate [95, 125, 214–216, 218]	? (n = 3); Moderate [217–219]	+(n = 1); Low [125]	B
	Primary care (n = 1) [221]									+(n = 1); Moderate [117, 221]	B
	Home or community (n = 1) [220]			? (n = 1); Moderate [220]		? (n = 1); High [220]		+(n = 1); low [220]			B
Doctors' Interpersonal Skills Questionnaire (DISQ)	Unidentified (n = 1) [222]										B
	Primary care (n = 2) [74, 223]	+(n = 1); Low [74]	? (n = 2); Moderate [74, 223]	? (n = 2); High [74, 223]				+(n = 1); High [74]	+(n = 1); Low [74]	+(n = 1); Very Low [74]	B
	Acute or Speciality (n = 3) [65, 114, 115]	+(n = 2); Moderate [65, 114]	? (n = 3); High [65, 114, 115]	? (n = 3); High [65, 114, 115]				+(n = 1); High [222]	+(n = 1); Very High [222]	+(n = 1); Low [222]	B
Interpersonal processes of care (IPC)	Primary care (n = 4) [116–119]	+(n = 1); Moderate [117]	? (n = 2); Moderate [116, 119]	? (n = 2); Moderate [116, 119]				+(n = 1); High [116]	+(n = 2); Very Low [65, 114]	+(n = 2); Low [114, 115]	A
	Home or community (n = 1) [113]	+(n = 1); Low [113]									B
	Public health (n = 1) [64]			? (n = 1); Moderate [64]		? (n = 1); High [64]					B

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Patient Participation Emergency Department questionnaire (pPED)	Acute or Speciality (n = 1) [87]	+(n = 1); Low [87]	Low	? (n = 1); Moderate [87]	? (n = 3); High [87]		+(n = 1); Low [87]			+(n = 1); Low [87]	B
	Acute or Speciality (n = 5) [224–228]	+(n = 3); Low [224, 225, 227]	Moderate [224, 227, 228]	? (n = 3); Moderate [224, 227, 228]	? (n = 5); Moderate [224–228]		+(n = 2); Very low [224, 225]	+(n = 1); Very low [226]		+(n = 2); Low [224, 225]	B
	Primary care (n = 1) [88]	+(n = 1); Low [88]	Low	? (n = 1); High [88]	? (n = 1); High [88]					? (n = 1); Low [88]	B
Role-Based Performance Scale	Other—Non-healthcare (n = 1) [89]	+(n = 1); Low [89]	Low	-(n = 1); Moderate [89]	? (n = 1); High [89]		+(n = 1); High [89]				B
Shared Decision-Making Questionnaire (SDM-Q-9)	Acute or Speciality (n = 17) [90, 120–134, 139]	+(n = 11); Low [90, 120, 122, 125, 127, 128, 130, 131, 133, 134, 139]	Low	+(n = 15); Moderate [90, 120–124, 127, 128, 130–134, 139]	+(n = 16); High [90, 120–125, 127, 127–134, 139]	+(n = 2); Moderate [121, 124–130, 133]	Moderate [90, 124–130, 133]	+(n = 9); Low [124, 128, 130, 133]		+(n = 2); Very Low [127, 129]	A
	Primary care (n = 6) [90, 135–139]	+(n = 5); Low [90, 135, 136, 138, 139]	Low	+(n = 6); Moderate [90, 135–139]	+(n = 5); Moderate [90, 135, 136, 138, 139]		+(n = 2); Moderate [90, 136]	+(n = 2); Low [136, 139]		-(n = 1); Low [137]	A
The Patient Assessment of Chronic Illness Care Survey (PACIC)	Acute or Speciality (n = 16) [140–150, 163, 166–168, 231]	+(n = 9); Moderate [144, 146–150, 166–168]	High [140–145, 147–150, 163, 166–168, 231]	? (n = 15); High [140–145, 147–150, 163, 166–168, 231]	+(n = 14); High [140, 142, 143, 145–150, 163, 166–168, 231]		+(n = 6); Moderate [141, 145, 149, 150, 163, 167]	? (n = 5); Low [143, 144, 147, 148, 167]		+(n = 1); ? (n = 5); Moderate-High ate [140, 141, [142] 143, 146, 168]	A
	Primary care (n = 16) [91, 151–160, 164, 169–171, 230]	+(n = 8); Low [91, 151, 152, 159, 160, 164, 170, 230]	Low	? (n = 15); Moderate [91, 91, 151–154, 151–160, 164, 169–171]	+(n = 15); High [91, 151–154, 156–160, 164, 169–171, 230]		+(n = 7); Moderate [91, 154, 157, 159, 160, 164, 169–171]	? (n = 5); Moderate [91, 152, 154, 159, 160, 164, 169–171]		? (n = 10); Moderate [91, 151, 152, 155, 157, 159, 160, 169–171]	A
Rehabilitation Unidentified (n = 1) [162]	Rehabilitation (n = 1) [161]			+(n = 1); High [161]	+(n = 1); High [161]		+(n = 1); Low [161]	+(n = 1); Low [161]			B
	Unidentified (n = 1) [162]			+(n = 1); High [162]	+(n = 1); High [162]						B

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
Urgent Care System Questionnaire (UCSQ)	Acute or speciality; Primary care; Long-term care; Home care; Home and community; and Public health (n = 1) [165]	+ (n = 1); Very low [165]	+ (n = 1); ate [165]	Model- + (n = 1); Low [165]							B
	Acute or speciality (n = 1) [92]	? (n = 1); Low [92]	? (n = 1); Moderate [92]	? (n = 1); High [92]					+ (n = 1); Low [92]		B
Ehrhart 2004 [75] Servant Leadership Scale	Home or Community (n = 1) [232]			+ (n = 1); Low [75]				+ (n = 1); Low [232]			B
	Non-healthcare – (n = 1) [75]			+ (n = 1); Low [75]	+ (n = 1); Low [75]			+ (n = 1); High [75]			B
	Acute or speciality (n = 1) [233]			? (n = 1); Low [233]	+ (n = 1); High [233]			+ (n = 1); High [233]	+ (n = 1); High [233]		B
Global Transformational Leadership Scale	Non-healthcare (n = 1) [76]			+ (n = 1); High [76]				+ (n = 1); High [76]			B
	Acute or speciality (n = 7) [77, 183–188]	? (n = 1); Low [76]	+ (n = 5); Low [77, 184–187]	+ (n = 6); High [77, 183, 185–188]	+ (n = 6); + (n = 1); Moderate [77, 183, ate [187]			+ (n = 5); High [77, 183, 185, 186, 188]	+ (n = 1); High [76]	+ (n = 4); Moderate [77, 183, 185, 187]	A
	Home or community (n = 2) [234, 235]			+ (n = 2); High [234, 235]	+ (n = 2); Low [234, 235]			+ (n = 1); High [235]		+ (n = 1); High [234]	B
Leadership Behavior Description Questionnaire [61]	No psychometric paper										
Servant Leadership Questionnaire	Non-healthcare (n = 1) [78]	+ (n = 1); Low [78]	+ (n = 1); High [78]	+ (n = 1); High [78]				+ (n = 1); Low [78]			A
Servant leadership Survey	Non-healthcare (n = 3) [79, 189, 190]	+ (n = 2); Low [79, 189]	+ (n = 2); High [79, 189, 190]	+ (n = 3); High [79, 189, 190]	+ (n = 3); + (n = 2); Moderate [79, 189, ate [79, 190]			+ (n = 3); High [79, 189, 190]	+ (n = 1); Low [190]	+ (n = 1); High [189]	A

Table 4 (continued)

Psychometric properties [rating (number of pertinent articles) COSMIN adapted GRADE quality of evidence]											
Instrument	Sector	CV	SV	IC	CCV/MI	ME	DCV	DKG	R	RL	COSMIN rating
The Mentorship Effectiveness Scale	Healthcare education (n = 1) [80]	? (n = 1); Very Low [80]									B
Organization of Work											
NASA Task Load Index	Acute or speciality (n = 3) [191, 192, 238]		High [191, 192, 238]	+ (n = 2); High [191, 192, 238]	+ (n = 3); High [191, 192, 238]		+ (n = 2); Moderate [191, 238]	+ (n = 1); Low [238]			B
	Healthcare education (n = 2) [237, 239]	? (n = 1); Low [237]	Moderate [239]	? (n = 1); Moderate [239]	? (n = 1); High [239]			? (n = 2); Very Low [237, 239]			B
	Non-healthcare (n = 1) [81]	+ (n = 1); Low [81]	Very low [81]	? (n = 1); Very low [81]						+ (n = 1); Very Low [81]	B
RAFAELA System: Professional Assessment of Nursing Care Intensity Level (PAONCIL) And Oulu Patient Classification Instrument (OPCq)	Acute or speciality (n = 9) [93, 193–197, 240–242]	? (n = 5); Low [93, 193, 197, 241, 242]		? (n = 1); High [193]	? (n = 1); High [193]		+ (n = 2); Low [93, 196]	+ (n = 3); Low [195–197]		? (n = 3); Very Low [193, 240, 242]	B
	Home or community (n = 1) [198]				? (n = 1); High [198]					- (n = 1); Low [198]	B

The measurement property ratings are rated as sufficient (+), insufficient (-), inconsistent (?), and not determined (?) and accompanied by the number of studies that evaluated the measurement property and an adapted GRADE measure of quality of evidence in brackets. The *n* values represent the number of studies for which evaluated a particular psychometric property

CV content validity, SV structural validity, IC internal consistency, CCV/MI cross-cultural validity/measurement invariance, ME measurement error, DKG discriminant validity, DCV discriminative/known groups validity, R responsiveness, RL reliability. Blank cells mean they were not assessed

¹Two studies [90, 139] evaluated the SDM-Q in both acute/specialty care and primary care settings; three studies investigated both the SDM-Q and the SDM-Q-Doc [127, 128, 137]

items) of the TCI in acute/specialty care [62, 63, 98], primary care [103], and public health [100] settings.

Physical and Technological Resources. Three instruments [58, 67, 68] assess *Physical and Technological Resources* (Table 4), defined as the physical structures and technological resources of an organization that are required to deliver services [25]. The instruments we mapped into this attribute were included because they measured technology, equipment, and supplies or measured online resources that are features of the physical and technological resources ICON attribute [25]. No psychometric papers were found for the Computer Systems and Decision-Making instrument [58]; thus, it could not be assessed to make a recommendation. The remaining two instruments [67, 68, 104] have the potential to be recommended for use; however, confirmatory factor analysis was not conducted to evaluate their theoretical factor structure, as required by the COSMIN guidelines. Therefore, further research is required.

Internal Relationships. Six instruments [69–71, 83, 94, 96] assess *Internal Relationships* (Table 4). Internal relationships are defined as “the ways in which two or more people or groups within an organization regard and behave toward each other” [25]. We attributed the six instruments to measuring the internal relationships because they pertained to partnerships or collaborative practices, which is a feature of internal relationship in the ICON framework [25]. All instruments have the potential to be recommended for use following further research. However, the Shared Decision-Making Questionnaire-Doc (SDM-Q-Doc) [71] can be recommended for use in long-term care, as sufficient (although low-quality) evidence of content validity and sufficient (moderate-quality) evidence of internal consistency were reported in two studies in long-term care [71, 105].

Organizational Culture. Three instruments [42, 84, 85] assess *Organizational Culture* (Table 4). Organizational culture is defined as “the normative beliefs and shared expectations that govern the work behaviour of an organization” [25]. We mapped the three instruments to organizational culture because they measured organizational norms, which is a feature of organizational culture in the ICON framework [25]. All three instruments had evidence of sufficient content validity (low quality) reported in a single article but lacked sufficient evidence of internal consistency to be recommended. An exploratory factor analysis was performed for two of the instruments [42, 84], and no internal consistency evidence was reported for the final instrument [85]. Therefore, all three instruments [42, 84, 85] have the potential to be recommended for use but first require more research.

Organizational Change Processes. One instrument, Measure of Champion Behaviour [106], assesses

Organizational Change Processes (Table 4). Organizational change processes is defined as “the process of altering an organization’s strategies, processes, procedures, technologies and/or culture to improve service delivery” [25]. We mapped Measure of Champion Behaviour [106] into the organizational change processes because champions/opinion leaders is a feature of this ICON attribute [25]. This instrument had sufficient evidence of high-quality content validity, structural validity, internal consistency, and convergent validity and is recommended for use. However, it was only evaluated in one psychometric study conducted in business settings, not in healthcare.

Receptivity to Change. Two instruments [73, 86] assess *Receptivity to Change* (Table 4) and are suitable for use in selected sectors. Receptivity to change is defined as “an openness and responsiveness to ideas, impressions or suggestions, and a readiness or fit of critical features of the environment as they specifically relate to a targeted practice” [25]. We mapped the two instruments to the receptivity to change attribute because they measured compatibility or readiness for change, which are features of this ICON attribute [25]. The Organizational Readiness for Implementing Change (ORIC) Scale [86] is recommended for use based on its development article, which reported assessments in multiple settings and sectors from acute care to university students and non-government organizations. However, in other studies, there is uncertainty regarding the evidence related to content validity of the ORIC in acute care ($n=2$ studies [107, 108]) and primary care ($n=1$ study [109]). In studies in the home/community sector [110] and primary care [109, 111], researchers found poor data fit with the two-factor structure theorized and tested in the development study [86].

The Resistance to Change Scale [73] is recommended for use based on one study [73] showing sufficient content validity (low quality) evidence and two studies [73, 112] identifying sufficient internal consistency and structural validity (high quality) evidence of a four-factor structure, but neither was conducted in healthcare. One of these studies demonstrated sufficient measurement invariance (high-quality) across data from 17 countries [112]. The other study showed sufficient discriminative validity between resistance to change and measures related to personality traits and cognitive ability and demonstrated the ability of the ‘resistance to change’ score to discriminate between individuals who are more likely to change behaviour after an imposed system change [73].

Evaluation Activity. Eleven instruments [59, 60, 64, 65, 74, 87–92, 95, 97] assess *Evaluation Activity* (Table 4), four of which are recommended for use: (1) Interpersonal Processes of Care (IPC) [64, 65]; (2) Shared Decision-Making Questionnaire (SDM-Q-9) [90]; (3) Patient

Assessment of Chronic Illness Care Survey (PACIC) [91]; (4) Care Transitions Measure (CTM) [97]. Evaluation activity is defined as “the systematic collection of information about the activities, characteristics and outcomes of programmes, services, policies or processes in an organization, to make judgements about the programme/process, improve effectiveness and/or inform decisions about future development in that organization” [25]. Although the instruments we mapped into the evaluation activity attribute may relate to other ICON attributes (e.g., measures of shared decision-making could also relate to internal relationships), the 11 instruments we included in the evaluation activity are intended to be used to garner patient/client/consumer feedback to staff regarding the services they received, which is a feature of the evaluation activity ICON attribute [25].

Different versions of the Interpersonal Processes of Care (IPC) instrument were developed based on the original conceptual framework and items [64]. Variants of the IPC I version (41–57 items) were evaluated in two studies, one in public health [64] and the other in a home/community setting [113]. Variants of the IPC II version (16–29 items) were evaluated in seven studies, three in acute/specialty care [65, 114, 115] and four in primary care [116–119]. The IPC II can be recommended for use in acute/specialty care [65, 114, 115] and in primary care [116–119], but there is insufficient evidence to recommend the IPC I for use in the home/community [113] and public health settings [64].

In a single study [90], the original SDM-Q questionnaire with 11 items was evaluated in both primary and acute care settings and is recommended for use in both settings. The remaining 21 studies evaluated a variant of the most updated version of the instrument, the SDM-Q-9 (7–9 items), in acute/specialty care ($n=16$ articles [120–134], primary care ($n=4$ articles [135–138], or both acute/specialty care and primary care ($n=1$ article [139]). The SDM-Q-9 is recommended for use in both acute/specialty care [120–134, 139] and primary care [120–134, 139].

The original PACIC (20 items) was evaluated in 24 studies (13–20 items) in acute or specialty care ($n=11$ articles [140–150]), primary care ($n=11$ articles [91, 151–160]), rehabilitation ($n=1$ article [161]), and an unidentified sector ($n=1$ article [162]). The short version of the PACIC (11 items) was evaluated in 6 studies in acute/specialty care ($n=4$ articles [142–144, 163]) and primary care ($n=2$ articles [151, 164]). In one study [165], the short version of the PACIC was evaluated with a mix of individuals from acute/specialty care, primary care, long-term care, home/community, and public health sectors. An extended version of the PACIC (PACIC5A), with six additional items that evaluate actions providers can

perform to facilitate behaviour change in patients, was evaluated in acute/specialty care ($n=3$ articles [166–168]) and primary care ($n=3$ articles [169–171]). Finally, a study conducted in primary care [153] evaluated a 27-item version of the PACIC, which included seven additional items assessing culturally sensitive care. The original PACIC is recommended for use in acute/specialty care [140–150] and in primary care [91, 151–160], whereas the short version of the PACIC is recommended for use in acute/specialty care [142–144, 163] and primary care [151, 164].

Seven studies [172–178] evaluated the psychometric properties of the CTM-3 in acute/specialty care, and 12 studies evaluated the psychometric properties of the CTM-15 in acute/specialty care ($n=10$ articles [172–181]), primary care ($n=1$ [97]), and rehabilitation ($n=1$ [182]). The CTM-3 and CTM-15 are recommended for use in acute/specialty care. However, it is important to note that despite the informed COSMIN recommendation, there was variability in the psychometric properties reported across the articles. For instance, while four articles [173–175, 179] reported that the original four-factor structure sufficiently fit the data, four other articles [177, 180, 181] did not.

Leadership. Seven instruments [61, 75–80] assess *Leadership* (Table 4), three of which are recommended for use: (1) the Implementation Leadership Scale in acute/specialty care ($n=7$ articles [77, 183–188]); (2) the Servant Leadership Questionnaire in a non-healthcare setting ($n=1$) [78]; (3) the Servant Leadership Survey also in non-healthcare settings ($n=3$ articles) [79, 189, 190]. Leadership “refers to the types and styles of leaders within an organization” [25]. We mapped six instruments to the leadership attribute because they measured leadership styles, which is a feature of the leadership attribute in the ICON framework [25]; the Mentorship Effectiveness Scale [80] pertained to the mentor feature of the leadership ICON attribute [25]. A lack of evidence regarding the content validity of the four remaining leadership instruments precludes recommending their use.

Organization of Work. Two instruments assess *Organization of Work* (Table 4): (1) NASA Task Load Index [81] and (2) RAFAELA System, which consists of the Professional Assessment of Nursing Care Intensity Level (PAONCIL) and Oulu Patient Classification instrument (OPC) [93]. Organization of work “refers to arrangement of tasks, responsibilities and resources within and between service providers working in the organization” [25]. We mapped the two instruments to the organization of work attribute because they measured workload, which is a feature of this ICON attribute [25]. We determined that both instruments have the potential to be recommended for use with more psychometric testing.

The 6-item unidimensional NASA Task Load Index was reported to have sufficient structural validity, internal consistency, and divergent and convergent validity when tested in diabetes management care [191] and in 18 intensive care units [192]. With the RAFAELA System, a study [193] conducted in a single hospital demonstrated the relevance of the OPC as a measure of nursing intensity based on the care required by patients, and another article [93] described the importance of using both an objective (OPC) and subjective (PAONCIL) measure of nursing intensity. In the other studies, researchers aimed to evaluate the minimum requirements (e.g., explanatory power and response rate) for the RAFAELA system to accurately measure nursing intensity [194], hypothesis testing between two hospitals [195], medical and surgical units [196], and liver versus kidney transplant patients [197]. Although the subscales of the OPC were reported to have $\alpha \geq 0.70$ in two studies [193, 198], no articles reported sufficient structural validity.

Pragmatic properties of the 38 instruments

Additional file 4 shows the detailed pragmatic assessment of the 38 instruments. Using the PAPERS criteria [43], none of the instruments were rated as excellent for all pragmatic properties pertaining to acceptability and easiness. However, all the instruments [42, 58–97] were rated as at least adequate for all four areas assessed, including those measuring acceptability (cost and language accessibility) and easiness (assessor burden for training and interpretation and length). Regarding accessibility, all instruments [42, 58–97] were found to be cost-free and/or available in the public domain and were rated as excellent using this criterion. Multiple instruments were rated to have adequate accessibility ($n=16$, 42.1%) [42, 58, 62, 63, 66–69, 72, 75–77, 80, 84, 89, 93, 96], with readability at the college level. Thirteen instruments (34.2%) were rated to have excellent accessibility [61, 64, 65, 73, 74, 78, 79, 82, 86, 88, 91, 92, 95, 97], with a readability level below the eighth grade. Finally, nine instruments (23.7%) were rated as having good language accessibility [59, 60, 70, 71, 81, 83, 85, 87, 90, 94], with a readability level between the 8th and 12th grades. Regarding *easiness*, 35 (92.1%) instruments required no specific training for their administration and were rated as excellent [42, 58–60, 64–92, 94–97]. The TCI [62, 63], Leadership Behavior Description Questionnaire [61], and RAFAELA system [93] were rated as good since the instruments included a free or inexpensive manual for self-training. All instruments except the RAFAELA system [93] ($n=37$, 97.4%) [42, 58–92, 94–97] were rated to have adequate assessor burden in terms of interpretation. Lastly, most instruments ($n=29$, 76.3%) had between 10 and 50 items [42, 58–65, 68–70, 72–75, 77–80, 82–89, 91, 92, 94, 96, 97],

which is considered a good length, and 9 instruments (23.7%) had excellent length with 10 or fewer items [66, 67, 71, 76, 81, 90, 93, 95].

Discussion

Summary of findings

ICON comprises three levels (macro, meso, and micro), six domains, and 22 attributes [25]. In this study, we reviewed instruments that measure a single ICON meso-level attribute and did not locate any instruments that mapped to ICON macro-level attributes. We identified 38 instruments that measured 9 of the 15 meso-level ICON framework attributes of context [25]. The number of instruments per attribute ranged from 1 (for the attribute *Organizational Change Processes*) to 11 (for the *Evaluation* attribute). Most instruments were: (1) developed in the United States ($n=21$, 55.3%); (2) written originally in English ($n=34$, 89.5%); (3) tested initially in acute care or specialty care settings ($n=18$, 47.4%), primarily in acute care hospitals ($n=15$, 39.4%); and (4) targeted healthcare service providers (23 of 38, 60.5%). Furthermore, we identified, assessed using the COSMIN guidelines, and reported on 169 studies with the primary aim of evaluating the psychometric properties of the 38 instruments. Most articles were published between 2012 and 2022.

Using COSMIN's psychometric recommendation ratings, 13 (34.2%) instruments measuring the following five ICON framework meso-level attributes met our criteria (could state them here) for recommending their use (Recommendation A—Table 4) in respective health sectors: For *Organizational Climate*: the (1) Internal Participation Scale in rehabilitation; (2) TCI in acute care/specialty care and primary care; (3) Measure of Trust in Teams in acute care/specialty care. For *Organizational Change Processes*: (1) Measure of Champion Behaviour in business settings. For *Receptivity to Change*: (1) ORIC scale in acute/specialty care and international non-government organizations and (2) the Resistance to Change Scale in academic settings. For *Evaluation Activity*: (1) the CTM in acute/specialty care; (2) the IPC in acute/specialty care and primary care; (3) the SDM-Q-9 in acute/specialty care and primary care; (4) the PACIC in acute/specialty care and primary care. For *Leadership*: (1) the Implementation Leadership Scale in acute/specialty care; (2) the Servant Leadership Questionnaire in non-healthcare settings; (3) the Servant Leadership Survey in non-healthcare settings. The remaining instruments with identified psychometric articles ($n=22$, 57.9%) have the potential to be recommended for use but first require further psychometric testing (Recommendation B—Table 4). Barriers to recommending these instruments currently relate to: (1) limited evidence of content validity, including: (a) the

instrument has not been subjected to an assessment of content validity; (b) lack of description of the sample and inclusion of the potential users of instruments during cognitive testing or pilot testing; (c) lack of description of the methods used to evaluate relevance, comprehensiveness, and comprehensibility; and (d) lack of detailed findings related to content validity); (2) lack of theory-informed evaluation of the instrument's structural validity (e.g., confirmatory factor analysis).

In terms of pragmatic properties, all 38 instruments meet some criteria for acceptability and easiness (Additional file 4).

Content validity

Borsboom et al. [199] and De Vet et al. [200] argue that content validity is the most important type of validity. De Vet et al. [200] recommend sequential evaluation of content validity, as instruments are applied within different contexts and/or for different purposes. Establishing content validity is primarily a qualitative process that involves: (1) determining a clear theoretical understanding of the construct/concept to be measured and the context in which it will be evaluated; (2) receiving input from a diverse expert panel (including potential instrument users) on the construct/concept; (3) assessing alignment of the instrument's content (including how to use it) with the construct/concept by the expert panel; and (4) using a framework to evaluate the correspondence between the instrument and the intended construct [200]. We agree that content validity is a critical psychometric property since, without it, other measurement properties of the instrument will be less meaningful. That said, it is not a simple form of validity to establish or assess. One reason for this, as illustrated in our review, is that instruments measuring the same construct or attribute often differ regarding the context for which they were developed. For example, the IPC [64, 65] (attribute—*Evaluation Activity*) was developed to measure the perspectives of equity-deserving populations regarding interpersonal processes of care (e.g., timeliness, communication, respectfulness) in acute/specialty care [65, 114, 115]. However, multiple variations have been developed and are recommended for use in primary care [116–119], with potential for use in home/community care [113] and public health [64] (Table 3). Further illustrating this complexity, we found evidence regarding the comprehensibility of items when instruments are re-evaluated for content validity after translation into a different language or used for the first time in a different sector or setting. In these instances, our COSMIN ratings for instrument development quality were often 'inadequate' or 'doubtful' because the authors did not adequately describe their process.

Other identified weaknesses in content validity for most of the 38 instruments across 169 psychometric papers involved the authors: (1) not reporting the use of cognitive interviews to evaluate the comprehensiveness or comprehensibility of the instruments, (2) inadequately describing the content experts who appraised the instruments, or (3) not including the intended users of the instruments in the content validation process. These shortcomings have been identified in recent systematic reviews [201–203] using COSMIN. In addition to content validity, another dimension that could be used to assess the appropriateness of a tool is the key informants' interpretation of the instrument's acceptability, relative advantage over existing measures, ease of completion, appropriateness to the context, fit with existing organizational activities, and the ability of the data garnered from the instrument to inform clinical or organizational decision-making [43]. From a Learning Health System perspective, engaging key informants in co-design will likely ensure instruments align with user needs, enhancing their relevance and effectiveness in improving health outcomes [204].

Limitations

Although rigorous methods were used for this review, there are four main study limitations. First, we needed to utilize a more explorative and practical process to locate instruments by limiting our Stage 1 search to review articles, considering the complexity of ICON. Hence, while we identified 38 instruments, which were assessed in 169 papers across multiple healthcare sectors and settings, it is possible that not all relevant instruments were captured in these existing reviews. Similarly, instruments measuring concepts like the instruments we included may have been missed in our search, as they may not have been detailed in the articles (primarily reviews on context) we used to find tools. Furthermore, we focused on instruments that assess a single context attribute at the meso- or macro-levels of ICON; hence, instruments that evaluate multiple attributes ($n=93$ instruments) were not included in our analysis. Third, we primarily searched for psychometric articles in a single database (PubMed) using the methodological filter developed by Terwee and colleagues [54]. While this may initially be perceived as a limitation, we found that the use of the precise methodological search filter in combination with reference checks was more sensitive than conducting a citation search [55]. Fourth, although we analyzed our findings according to different sectors, which aligns with the idea that context can influence psychometric properties, the COSMIN manual [56, 57] does not currently provide guidance on analyzing data according to different sectors and settings.

Conclusion

In summary, we located 38 instruments that measure ICON meso-level attributes of context [25] and assessed 169 psychometric papers on these instruments based on the COSMIN guidelines [56, 57]. Only 13 instruments met our criteria for recommending their use. We will be conducting a second psychometric review of instruments identified in the micro (individual) level of ICON and will then develop a comprehensive and pragmatic context survey that targets all 22 ICON attributes. Understanding and measuring context is crucial for the successful implementation and operation of Learning Health Systems, which rely on continuous improvement informed by contextual factors. Future work should also include: (1) additional psychometric assessments of the 25 instruments not yet recommended for use based on our analysis, given inadequate existing evidence, (2) psychometric and pragmatic assessments of instruments that assess multiple attributes at the meso- or macro-levels of ICON; (3) the development of additional instruments to target the ICON attributes important for successful implementation and for which instruments were not identified ($n=6$ meso-level and $n=5$ macro-level), and (4) the development of an open access registry of the instruments identified in this review and a website to facilitate access to the instruments and their psychometric and pragmatic properties.

Abbreviations

ABIM	American Board of Internal Medicine's questionnaire
CTM	Care Transitions Measure
CIROP	Community Impacts of Research Oriented Partnerships
COSMIN	COnsensus-based Standards for the selection of health Measurement Instruments
DISQ	Doctors' Interpersonal Skills Questionnaire
ICON	Implementation in Context
IPC	Interpersonal Processes of Care
MECIR	Methodological Expectations of Cochrane Intervention Reviews
ORIC	Organizational Readiness for Implementing Change
OPC	Oulu Patient Classification instrument
OMP	Outcomes Measurement Properties
PACIC	Patient Assessment of Chronic Illness Care Survey
PPED	Patient Participation Emergency Department questionnaire
PROMs	Patient Reported Outcomes Measures
PHWSUQ	Perceived Health Web Site Usability Questionnaire
PAONCIL	Professional Assessment of Nursing Care Intensity Level
SDM-Q-9	Shared Decision-Making Questionnaire
SDM-Q-Doc	Shared Decision-Making Questionnaire-Doc
TCI	Team Climate Inventory
UCSQ	Urgent Care System Questionnaire

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12961-026-01507-z>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Acknowledgements

We thank Danielle Starr (DS) and Junqiang Zhao (JZ) for their contributions in the early phase of data collection. The ICON team: Chantal Backman¹, Anna Bergström², Jamie Brehaut^{3,4}, Melissa Brouwers⁴, Christopher Burton⁵, Christine Cassidy⁶, Cheyne Chalmers⁷, Anna Chapman^{7,8}, Heather Colquhoun⁹, Janet Curran⁶, Melissa Demery Varin¹, Paula Doering¹⁰, Annette Elliott Rose¹¹, Christina Godfrey¹², Doris Grinspun¹³, Gillian Harvey¹⁴, Noah Ivers^{15,16}, Shelly-Anne Li¹⁷, Letitia Nadalin-Penno¹⁸, Tamara Rader¹⁹, Jo Rycroft-Malone²⁰, Dawn Stacey^{1,3}, Sharon Straus^{21,22,23}, Andrea C. Tricco^{21,24}, Lars Wallin²⁵, Chantal Backman²⁶

¹University of Ottawa, School of Nursing, Ottawa, Canada, ²Uppsala University, Department of Women's and Children's Health, Uppsala, Sweden, ³Ottawa Hospital Research Institute, Methods and Implementation Research Program, Ottawa, Canada, ⁴University of Ottawa, School of Epidemiology and Public Health, Ottawa, Canada, ⁵University of East Anglia, School of Health Science, Norwich, United Kingdom, ⁶Dalhousie University, School of Nursing, Halifax, Canada, ⁷Deakin University, School of Nursing and Midwifery, Geelong, Australia, ⁸Deakin University, Centre for Quality and Patient Safety Research, Institute for Health Transformation, Geelong, Australia, ⁹University of Toronto, Department of Occupational Science and Occupational Therapy, Toronto, Canada, ¹⁰Élisabeth Bruyère Health, Ottawa, Canada, ¹¹Nova Scotia Health Authority, Nova Scotia, Canada, ¹²Queen's University, School of Nursing, Kingston, Canada, ¹³Registered Nurses' Association of Ontario (RNAO), Toronto, Canada, ¹⁴Flinders University, College of Nursing and Health Sciences, Bedford Park, South Australia, Australia, ¹⁵University of Toronto, Department of Family and Community Medicine, Toronto, Canada, ¹⁶Women's College Hospital Research and Innovation Institute, Toronto, Canada, ¹⁷Artists' Health Centre, Department of Family and Community Medicine, University Health Network, Toronto, Canada, ¹⁸Faculty of Environmental Studies, Biotechnology and Health Sciences, Canadore College North Bay, Toronto, ¹⁹Canadian Agency for Drugs and Technologies in Health, Ottawa, Canada, ²⁰Lancaster University, Faculty of Health and Medicine, Lancaster, United Kingdom, ²¹Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Ontario, Canada, ²²University of Toronto, Department of Medicine, Toronto, Ontario, Canada, ²³University of Toronto, Dalla Lana School of Public Health, Toronto, Ontario, Canada, ²⁴Epidemiology Division & Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada, ²⁵Dalarna University, School of Health and Welfare, Falun, Sweden, ²⁶University of Ottawa, School of Nursing, Ottawa, Canada

Author contributions

All the authors participated in the conception of the study. All authors contributed to the development of the study design. TR developed and ran the search strategy used to identify instruments included in the review (Stage 1). DS, JZ, LA, MDV, and WJS conducted screening and data extraction for Stage 1. AMH, IDG, JES, JZ, LA, LKC, MDV, and WJS mapped each individual item of an included instrument to an ICON framework. WJS ran the search for Stage 2 (psychometric review of included instruments). LKC, MG, VJW, and WJS conducted screening and data extraction for Stage 2. AMH, IDG, JES, LKC, MG, VJW, and WJS completed synthesis of the COSMIN and pragmatic assessments. AMH, IDG, JES, and WJS drafted the manuscript. All team members critically reviewed and revised the manuscript regarding its content, approved the version to be published, and agreed to be accountable for all aspects of the work.

Funding

Canadian Institutes for Health Research; Grant number: 426122.

Data availability

The search strategy, the list of excluded articles, and pragmatic assessments are provided as additional files. The datasets used and/or analyzed during

the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Health Sciences, School of Nursing, University of Ottawa, Ottawa, Canada. ²Ottawa Hospital Research Institute, Methods and Implementation Research Program, Ottawa, Canada. ³Faculty of Medicine, School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada. ⁴School of Nursing and Midwifery, Deakin University, Centre for Quality and Patient Safety Research in the Institute for Health Transformation, Geelong, Australia. ⁵Children's Hospital of Eastern Ontario, Ottawa, Canada. ⁶Barwon Health, Geelong, Australia.

Received: 30 January 2026 Accepted: 17 June 2026

Published online: 28 July 2026

References

- Grinspun D, Mazurek Melnyk B, Fineout-Overholt E, Naik S, Wallace K. Advancing optimal care with robust clinical practice guidelines. Evidence-Based Practice in Nursing & Healthcare: A Guide to Best Practice. 5th ed. 2023.
- Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care*. 2001;39(8 Suppl 2):146–54.
- Lang ES, Wyer PC, Haynes RB. Knowledge translation: closing the evidence-to-practice gap. *Ann Emerg Med*. 2007;49(3):355–63.
- Lauer MS, Skarlatos S. Translational research for cardiovascular diseases at the National Heart, Lung, and Blood Institute: moving from bench to bedside and from bedside to community. *Circulation*. 2010;121(7):929–33.
- McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635–45.
- Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? 1998. *Milbank Q*. 2005;83(4):843–95.
- Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JL, Coiera EW, et al. CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust*. 2012;197(2):100–5.
- Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. *Qual Health Care*. 2001;10(3):152–8.
- Squires JE, Cho-Young D, Aloisio LD, Bell R, Bornstein S, Brien SE, et al. Inappropriate use of clinical practices in Canada: a systematic review. *CMAJ*. 2022;194(8):E279–96.
- Pomare C, Mahmoud Z, Vedovi A, Ellis LA, Knaggs G, Smith CL, et al. Learning health systems: a review of key topic areas and bibliometric trends. *Learn Health Syst*. 2022;6(1):e10265.
- Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implement Sci*. 2015;10(1):155.
- Reichenpfader U, Carlford S, Nilsen P. Leadership in evidence-based practice: a systematic review. *Leadersh Health Serv*. 2015;28(4):298–316.
- Wozney N, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. E-Mental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. *JMIR Ment Health*. 2018. <https://doi.org/10.2196/mental.9655>.
- Li S-A, Jeffs L, Barwick M, Stevens B. Organizational contextual features that influence the implementation of evidence-based practices across healthcare settings: a systematic integrative review. *Syst Rev*. 2018;7(1):72.
- De Vries H, Bekkers V, Tummers L. Innovation in the public sector: a systematic review and future research agenda. *Public Adm*. 2016;94(1):146–66.
- Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, et al. Achieving change in primary care—causes of the evidence to practice gap: systematic reviews of reviews. *Implement Sci*. 2015;11(1):40.
- May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res*. 2007;7:148.
- Coles E, Wells M, Maxwell M, Harris FM, Anderson J, Gray NM, et al. The influence of contextual factors on healthcare quality improvement initiatives: what works, for whom and in what setting? Protocol for a realist review. *Syst Rev*. 2017;6(1):168.
- Taylor SL, Dy S, Foy R, Hempel S, McDonald KM, Øvretveit J, et al. What context features might be important determinants of the effectiveness of patient safety practice interventions? *BMJ Qual Saf*. 2011;20(7):611–7.
- Bird M, MacPhee M, Shaw J, Wodchis WP, Jeffs L, Austin T, et al. Evaluating for learning and sustainability (ELS) framework: a realist synthesis. *BMC Health Serv Res*. 2025;25(1):683.
- Squires JE, Graham I, Bashir K, Nadalin-Penno L, Lavis J, Francis J, et al. Understanding context: a concept analysis. *J Adv Nurs*. 2019;75(12):3448–70.
- Squires JE, Graham ID, Hutchinson AM, Linklater S, Brehaut JC, Curran J, et al. Understanding context in knowledge translation: a concept analysis study protocol. *J Adv Nurs*. 2015;71(5):1146–55.
- Squires JE, Aloisio LD, Grimshaw JM, Bashir K, Dorrance K, Coughlin M, et al. Attributes of context relevant to healthcare professionals' use of research evidence in clinical practice: a multi-study analysis. *Implement Sci*. 2019;14(1):52.
- Squires JE, Hutchinson AM, Coughlin M, Bashir K, Curran J, Grimshaw JM, et al. Stakeholder perspectives of attributes and features of context relevant to knowledge translation in health settings: a multi-country analysis. *Int J Health Policy Manag*. 2022;11(8):1373–90.
- Squires JE, Graham ID, Santos WJ, Hutchinson AM. The Implementation in Context (ICON) framework: a meta-framework of context domains, attributes and features in healthcare. *Health Res Policy Syst*. 2023;21(1):81.
- Packiam P, Graham ID, Graham N, Hutchinson A, Santos WJ, Squires JE. Assessing the applicability of the Implementation in CONtext (ICON) framework to low and middle income countries: A secondary analysis (In Review). *Health Res Policy Syst*. 2026.
- Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. *Implement Sci*. 2015;11(1):1–13.
- Kitson A, Harvey G, McCormack B. Enabling the implementation of evidence based practice: a conceptual framework. *Qual Health Care*. 1998;7(3):149–58.
- Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Health Prof*. 2006;26(1):13–24.
- Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice. A consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50.
- Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. *Implement Sci*. 2022;17(1):1–16.
- Squires JE, Hutchinson AM, Coughlin M, Bashir K, Curran J, Grimshaw JM, et al. Stakeholder perspectives of attributes and features of context relevant to knowledge translation in health settings: a multi-country analysis. *Int J Health Policy*. 2022;11(8):1373–90.
- Chaudoir SR, Dugan AG, Barr CH. Measuring factors affecting implementation of health innovations: a systematic review of structural,

- organizational, provider, patient, and innovation level measures. *Implement Sci.* 2013;8:22.
34. Brennan SE, Bosch M, Buchan H, Green SE. Measuring organizational and individual factors thought to influence the success of quality improvement in primary care: a systematic review of instruments. *Implement Sci.* 2012;7(1):121.
 35. Rogers L, De Brún A, McAuliffe E. Defining and assessing context in healthcare implementation studies: a systematic review. *BMC Health Serv Res.* 2020;20(1):591.
 36. Society for Implementation Research Collaboration (SIRC). The SIRC Instrument Review Project (IRP): Society for Implementation Research Collaboration (SIRC); 2020. <https://societyforimplementationresearchcollaboration.org/sirc-instrument-project/>.
 37. McHugh S, Dorsey CN, Mettert K, Purtle J, Bruns E, Lewis CC. Measures of outer setting constructs for implementation research: a systematic review and analysis of psychometric quality. *Implement Res Pract.* 2020;1:2633489520940022.
 38. Powell BJ, Mettert KD, Dorsey CN, Weiner BJ, Stanick CF, Lengnick-Hall R, et al. Measures of organizational culture, organizational climate, and implementation climate in behavioral health: a systematic review. *Implement Res Pract.* 2021;2:26334895211018864.
 39. Dorsey CN, Mettert KD, Puspitasari AJ, Damschroder LJ, Lewis CC. A systematic review of measures of implementation players and processes: summarizing the dearth of psychometric evidence. *Implement Res Pract.* 2021;2:26334895211002470.
 40. Weiner BJ, Mettert KD, Dorsey CN, Nolen EA, Stanick C, Powell BJ, et al. Measuring readiness for implementation: a systematic review of measures' psychometric and pragmatic properties. *Implement Res Pract.* 2020;1:2633489520933896.
 41. Stanick C, Halko H, Mettert K, Dorsey C, Moullin J, Weiner B, et al. Measuring characteristics of individuals: an updated systematic review of instruments' psychometric properties. *Implement Res Pract.* 2021;2:26334895211000456.
 42. Connelly LM, Powers JL. Advances in patient safety online patient safety climate survey: tool development and lessons learned. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. *Advances in patient safety: from research to implementation*. Rockville: Agency for Healthcare Research and Quality (US); 2005.
 43. Stanick CF, Halko HM, Nolen EA, Powell BJ, Dorsey CN, Mettert KD, et al. Pragmatic measures for implementation research: development of the psychometric and pragmatic evidence rating scale. *Transl Behav Med.* 2021;11(1):11–20.
 44. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539–49.
 45. Terwee CB, Prinsen CA, Chiarotto A, Westerman M, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159–70.
 46. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* 2018;169(7):467–73.
 47. Elsmann EBM, Mokkink LB, Terwee CB, Beaton D, Gagnier JJ, Tricco AC, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *J Clin Epidemiol.* 2024;173:111422.
 48. Squires J, Hutchinson A, Graham I, Santos W, team I. Review of Tools that Could be Used to Measure Context in Healthcare: OSF; 2022. <https://osf.io/b2z8k>.
 49. McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evid Based Libr Inf Pract.* 2010;5(1):149–54.
 50. McMaster University Health Information Research Unit. Search Filters For Medline In Ovid Syntax And The Pubmed Translation: McMaster University 2022. <https://hiruweb.mcmaster.ca/hkr/hedges/medline/>.
 51. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. *J Med Libr Assoc.* 2018;106(4):531–41.
 52. Clarivate. EndNote 2023. <https://endnote.com/?language=en>.
 53. DistillerSR. DistillerSR: DistillerSR; 2023. <https://www.distillersr.com/products/distillersr-systematic-review-software>.
 54. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18(8):1115–23.
 55. Santos WJ, Hutchinson AM, Rader T, Graham ID, Watkins V, Candido LKd, et al. Insights from using an outcomes measurement properties search filter and conducting citation searches to locate psychometric articles of tools used to measure context attributes. *BMC Res Notes.* 2023;16(1):34.
 56. Terwee CB, Prinsen C, Chiarotto A, De Vet H, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs—user manual 2018. <https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>.
 57. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, De Vet H, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User Manual. 2018;78(1):6–3.
 58. Hatcher M. Survey of acute care hospitals in the United States relative to technology usage and technology transfer. *J Med Syst.* 1997;21:323–36.
 59. Webster G, Swanson D. Final report on the patient satisfaction questionnaire project. Philadelphia: American Board of Internal Medicine; 1989.
 60. Stump TE, Dexter PR, Tierney WM, Wolinsky FD. Measuring patient satisfaction with physicians among older and diseased adults in a primary care municipal outpatient setting: an examination of three instruments. *Med Care.* 1995;33(9):973–82.
 61. Stogdill RM. *Handbook of leadership: a survey of theory and research*. Free Press; 1974.
 62. Anderson N, West MA. The Team Climate Inventory: development of the TCI and its applications in teambuilding for innovativeness. *Eur J Work Organ Psychol.* 1996;5(1):53–66.
 63. Anderson NR, West MA. Measuring climate for work group innovation: development and validation of the team climate inventory. *J Organ Behav.* 1998;19(3):235–58.
 64. Stewart AL, Nápoles-Springer A, Pérez-Stable EJ. Interpersonal processes of care in diverse populations. *Milbank Q.* 1999;77(3):305–39.
 65. Stewart AL, Nápoles-Springer AM, Gregorich SE, Santoyo-Olsson J. Interpersonal processes of care survey: patient-reported measures for diverse groups. *Health Serv Res.* 2007;42(3 Pt 1):1235–56.
 66. Körner M, Wirtz MA. Development and psychometric properties of a scale for measuring internal participation from a patient and health care professional perspective. *BMC Health Serv Res.* 2013;13:374.
 67. Nahm ES, Resnick B, Mills ME. Development and pilot-testing of the perceived health Web Site usability questionnaire (PHWSUQ) for older adults. *Stud Health Technol Inform.* 2006;122:38–43.
 68. Murphy CA, Maynard M, Morgan G. Pretest and post-test attitudes of nursing personnel toward a patient care information system. *Comput Nurs.* 1994;12(5):239–44.
 69. Weiss SJ, Davis HP. Validity and reliability of the Collaborative Practice Scales. *Nurs Res.* 1985;34(5):299–305.
 70. Kenaszchuk C, Reeves S, Nicholas D, Zwarenstein M. Validity and reliability of a multiple-group measurement scale for interprofessional collaboration. *BMC Health Serv Res.* 2010;10:83.
 71. Scholl I, Kriston L, Dirmaier J, Buchholz A, Härter M. Development and psychometric properties of the Shared Decision Making Questionnaire—physician version (SDM-Q-Doc). *Patient Educ Couns.* 2012;88(2):284–90.
 72. Howell JM, Shea CM, Higgins CA. Champions of product innovations: defining, developing, and validating a measure of champion behavior. *J Bus Venturing.* 2005;20(5):641–61.
 73. Oreg S. Resistance to change: developing an individual differences measure. *J Appl Psychol.* 2003;88(4):680–93.
 74. Greco M, Cavanagh M, Brownlea A, McGovern J. The Doctor's Interpersonal Skills Questionnaire (DISQ): a validated instrument for use in GP training. *Educ Gen Pract.* 1999;10:256–64.
 75. Ehrhart M, Naumann S. Leadership and procedural justice climate citizenship behavior as antecedents of unit-level organizational. *Pers Psychol.* 2004;57(1):61–94.

76. Carless SA, Wearing AJ, Mann L. A short measure of transformational leadership. *J Bus Psychol*. 2000;14:389–405.
77. Aarons GA, Ehrhart MG, Farahnak LR. The Implementation Leadership Scale (ILS): development of a brief measure of unit level implementation leadership. *Implement Sci*. 2014;9(1):45.
78. Barbutto JJE, Wheeler DW. Scale development and construct clarification of servant leadership. *Group Organ Manag*. 2006;31(3):300–26.
79. van Dierendonck D, Nuijten I. The servant leadership survey: development and validation of a multidimensional measure. *J Bus Psychol*. 2010;26(3):249–67.
80. Berk RA, Berg J, Mortimer R, Walton-Moss B, Yeo TP. Measuring the effectiveness of faculty mentoring relationships. *Acad Med*. 2005;80(1):66–71.
81. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, 1988. *Advances in Psychology*. 1988.
82. Costa AC, Anderson N. Measuring trust in teams: development and validation of a multifaceted measure of formative and reflective indicators of team trust. *Eur J Work Organ Psychol*. 2011;20(1):119–54.
83. Jones J, Barry MM. Developing a scale to measure trust in health promotion partnerships. *Health Promot Int*. 2011;26(4):484–91.
84. Singer S, Meterko M, Baker L, Gaba D, Falwell A, Rosen A. Workforce perceptions of hospital safety culture: development and validation of the patient safety climate in healthcare organizations survey. *Health Serv Res*. 2007;42(5):1999–2021.
85. Schutz AL, Counte MA, Meurer S. Development of a patient safety culture measurement tool for ambulatory health care settings: analysis of content validity. *Health Care Manage Sci*. 2007;10(2):139–49.
86. Shea CM, Jacobs SR, Esserman DA, Bruce K, Weiner BJ. Organizational readiness for implementing change: a psychometric assessment of a new measure. *Implement Sci*. 2014;9:7.
87. Frank C, Asp M, Fridlund B, Baigi A. Questionnaire for patient participation in emergency departments: development and psychometric testing. *J Adv Nurs*. 2011;67(3):643–51.
88. Risser NL. Development of an instrument to measure patient satisfaction with nurses and nursing care in primary care settings. *Nurs Res*. 1975;24(1):45–52.
89. Welbourne TM, Johnson DE, Erez A. The role-based performance scale: validity analysis of a theory-based measure. *Acad Manage J*. 1998;41(5):540–55.
90. Simon D, Schorr G, Wirtz M, Vodermaier A, Caspari C, Neuner B, et al. Development and first validation of the shared decision-making questionnaire (SDM-Q). *Patient Educ Couns*. 2006;63(3):319–27.
91. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). *Med Care*. 2005;43(5):436–44.
92. O'Cathain A, Knowles E, Nicholl J. Measuring patients' experiences and views of the emergency and urgent care system: psychometric testing of the urgent care system questionnaire. *BMJ Qual Saf*. 2011;20(2):134–40.
93. Fagerström L, Rainio AK. Professional assessment of optimal nursing care intensity level: a new method of assessing personnel resources for nursing care. *J Clin Nurs*. 1999;8(4):369–79.
94. King G, Servais M, Kertoy M, Specht J, Currie M, Rosenbaum P, et al. A measure of community members' perceptions of the impacts of research partnerships in health and social services. *Eval Program Plann*. 2009;32(3):289–99.
95. Elwyn G, Barr PJ, Grande SW, Thompson R, Walsh T, Ozanne EM. Developing CollaboRATE: a fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient Educ Couns*. 2013;93(1):102–7.
96. Batorowicz B, Shepherd TA. Measuring the quality of transdisciplinary teams. *J Interprof Care*. 2008;22(6):612–20.
97. Coleman EA, Smith JD, Frank JC, Eilertsen TB, Thiare JN, Kramer AM. Development and testing of a measure designed to assess the quality of care transitions. *Int J Integr Care*. 2002;2:e02.
98. Ouwens M, Hulscher M, Akkermans R, Hermens R, Grol R, Woltersheim H. The Team Climate Inventory: application in hospital teams and methodological considerations. *Qual Saf Health Care*. 2008;17(4):275–80.
99. Kivimäki M, Elovainio M. A short version of the Team Climate Inventory: development and psychometric properties. *J Occup Organ Psychol*. 1999;72(2):241–6.
100. Kivimäki M, Kuk G, Elovainio M, Thomson L, Kalliomäki-Levanto T, Heikkilä A. The team climate inventory (TCI)- four or five factors? Testing the structure of TCI in samples of low and high complexity jobs. *J Occup Psychol*. 1997;70:375–89.
101. Beaulieu MD, Dragieva N, Del Grande C, Dawson J, Haggerty JL, Barnsley J, et al. The team climate inventory as a measure of primary care teams' processes: validation of the French version. *Healthcare Policy*. 2014;9(3):40–54.
102. Strating MM, Nieboer AP. Psychometric test of the Team Climate Inventory-short version investigated in Dutch quality improvement teams. *BMC Health Serv Res*. 2009;9(1):1–7.
103. Silva MC, Peduzzi M, Sangaletti CT, Silva D, Agreli HF, West MA, et al. Cross-cultural adaptation and validation of the teamwork climate scale. *Rev Saude Publica*. 2016;50:52.
104. Son H, Nahm ES, Zhu S, Galik E, Seidl KL, Van de Castle B, et al. Testing a model of patient portal use in adult patients. *J Nurs Scholarsh*. 2021;53(2):143–53.
105. Goto Y, Yamaguchi Y, Onishi J, Arai H, Härter M, Scholl I, et al. Adapting the patient and physician versions of the 9-item shared decision making questionnaire for other healthcare providers in Japan. *BMC Med Inform Decis Mak*. 2021;21(1):314.
106. Howell JM, Shea CM, Higgins CA. Champions of product innovations: defining, developing, and validating a measure of champion behavior. *J Bus Ventur*. 2005;20(5):641–61.
107. Lindig A, Hahlweg P, Christalle E, Scholl I. Translation and psychometric evaluation of the German version of the Organisational Readiness for Implementing Change measure (ORIC): a cross-sectional study. *BMJ Open*. 2020;10(6):e034380.
108. Storkholm MH, Mazzocato P, Tessma MK, Savage C. Assessing the reliability and validity of the Danish version of Organizational Readiness for Implementing Change (ORIC). *Implement Sci*. 2018;13(1):78.
109. Bomfim RA, Braff EC, Frazão P. Cross-cultural adaptation and psychometric properties of the Brazilian-Portuguese version of the Organizational Readiness for Implementing Change questionnaire. *Rev Bras Epidemiol*. 2020;23:e200100.
110. Ruest M, Léonard G, Thomas A, Desrosiers J, Guay M. French cross-cultural adaptation of the Organizational Readiness for Implementing Change (ORIC). *BMC Health Serv Res*. 2019;19(1):535.
111. Leslie HH, West R, Twine R, Masilela N, Steward WT, Kahn K, et al. Measuring organizational readiness for implementing change in primary care facilities in Rural Bushbuckridge, South Africa. *Int J Health Policy Manag*. 2022;11(7):912–8.
112. Oreg S, Bayazit M, Vakola M, Arciniega L, Armenakis A, Barkauskiene R, et al. Dispositional resistance to change: measurement equivalence and the link to personal values across 17 nations. *J Appl Psychol*. 2008;93(4):935.
113. Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, Stewart AL. Using cognitive interviews to develop surveys in diverse populations. *Med Care*. 2006;44(11 Suppl 3):S21–30.
114. Peimani M, Nasli-Esfahani E, Stewart AL, Shakibazadeh E, Yekaninejad MS, Garmaroudi G. Patient-physician communication in the context of diabetes care: adaptation and assessment of psychometric properties of a Persian version of the interpersonal process of care survey (P-IPC). *Diabetes Res Clin Pract*. 2020;162:108109.
115. Shiozawa T, Yamaguchi S, Matsunaga A, Sawada U, Fujii C. Development of the interpersonal processes of care survey-Japanese version. *Neuropsychopharmacol Rep*. 2020;40(1):107–12.
116. Beaulieu MD, Haggerty JL, Beaulieu C, Bouharaoui F, Lévesque JF, Pineault R, et al. Interpersonal communication from the patient perspective: comparison of primary healthcare evaluation instruments. *Healthc Policy*. 2011;7(Spec Issue):108–23.
117. Haggerty JL, Beaulieu C, Lawson B, Santor DA, Fournier M, Burge F. What patients tell us about primary healthcare evaluation instruments: response formats, bad questions and missing pieces. *Healthc Policy*. 2011;7(Spec Issue):66–78.
118. Haggerty JL, Bouharaoui F, Santor DA. Differential item functioning in primary healthcare evaluation instruments by French/English

- version, educational level and urban/rural location. *Healthcare Policy*. 2011;7(Spec Issue):47–65.
119. Haggerty JL, Burge F, Beaulieu MD, Pineault R, Beaulieu C, Lévesque JF, et al. Validation of instruments to evaluate primary healthcare from the patient perspective: overview of the method. *Healthcare Policy*. 2011;7(Spec Issue):31–46.
 120. Alvarado-Villa GE, Moncayo-Rizzo JD, Gallardo-Rumbea JA. Spanish validation endorsement of SDM-Q-9, a new approach. *BMC Public Health*. 2019;19(1):106.
 121. Alvarez K, Wang Y, Alegria M, Ault-Brutus A, Ramanayake N, Yeh YH, et al. Psychometrics of shared decision making and communication as patient centered measures for two language groups. *Psychol Assess*. 2016;28(9):1074–86.
 122. Baicus C, Balanescu P, Gurghean A, Badea CG, Padureanu V, Rezus C, et al. Romanian version of SDM-Q-9 validation in internal medicine and cardiology setting: a multicentric cross-sectional study. *Rom J Intern Med*. 2019;57(2):195–200.
 123. Ballesteros J, Moral E, Brieva L, Ruiz-Beato E, Prefasi D, Maurino J. Psychometric properties of the SDM-Q-9 questionnaire for shared decision-making in multiple sclerosis: item response theory modelling and confirmatory factor analysis. *Health Qual Life Outcomes*. 2017;15(1):79.
 124. Calderon C, Jiménez-Fonseca P, Ferrando PJ, Jara C, Lorenzo-Seva U, Beato C, et al. Psychometric properties of the shared decision-making questionnaire (SDM-Q-9) in oncology practice. *Int J Clin Health Psychol*. 2018;18(2):143–51.
 125. Hurley EA, Bradley-Ewing A, Bickford C, Lee BR, Myers AL, Newland JG, et al. Measuring shared decision-making in the pediatric outpatient setting: psychometric performance of the SDM-Q-9 and Collaborate among English and Spanish speaking parents in the US Midwest. *Patient Educ Couns*. 2019;102(4):742–8.
 126. Kasper J, Heesen C, Köpke S, Fulcher G, Geiger F. Patients' and observers' perceptions of involvement differ. Validation study on inter-relating measures for shared decision making. *PLoS ONE*. 2011;6(10):e26255.
 127. Nejati B, Lin CC, Imani V, Browall M, Lin CY, Broström A, et al. Validating patient and physician versions of the shared decision making questionnaire in oncology setting. *Health Promot Perspect*. 2019;9(2):105–14.
 128. Rodenburg-Vandenbussche S, Pieterse AH, Kroonenberg PM, Scholl I, van der Weijden T, Luyten GP, et al. Dutch translation and psychometric testing of the 9-item shared decision making questionnaire (SDM-Q-9) and shared decision making questionnaire-physician version (SDM-Q-Doc) in primary and secondary care. *PLoS ONE*. 2015;10(7):e0132158.
 129. Scholl I, Kriston L, Dirmaier J, Härter M. Comparing the nine-item shared decision-making questionnaire to the OPTION scale - an attempt to establish convergent validity. *Health Expect*. 2015;18(1):137–50.
 130. Schulz AA, Wirtz MA. Assessment of the quality of woman-centred midwifery care from the mothers' perspective: a structural analysis of cross-sectional survey data. *Z Evid Fortbild Qual Gesundhw*. 2021;166:8–17.
 131. Shunmugam B, Ng CJ, Aishah Mohd Taib N, Chinna K. Validation of the Malay, English, and Chinese translations of the 9-item shared decision making questionnaire (SDM-Q-9) in breast cancer patients making treatment decisions. *Asia Pac J Public Health*. 2021;33(8):914–22.
 132. Wu TY, Chen CT, Huang YJ, Hou WH, Wang JD, Hsieh CL. Rasch analysis of the 9-item shared decision making questionnaire in women with breast cancer. *Cancer Nurs*. 2019;42(3):E34–e42.
 133. Zisman-Ilani Y, Roe D, Scholl I, Härter M, Karnieli-Miller O. Shared decision making during active psychiatric hospitalization: assessment and psychometric properties. *Health Commun*. 2017;32(1):126–30.
 134. Alzubaidi H, Hussein A, Mc Namara K, Scholl I. Psychometric properties of the Arabic version of the 9-item shared decision-making questionnaire: the entire process from translation to validation. *BMJ Open*. 2019;9(4):e026672.
 135. De las Cuevas C, Perestelo-Perez L, Rivero-Santana A, Cebolla-Martí A, Scholl I, Härter M. Validation of the Spanish version of the 9-item shared decision-making questionnaire. *Health Expect*. 2015;18(6):2143–53.
 136. Goto Y, Miura H, Son D, Arai H, Kriston L, Scholl I, et al. Psychometric evaluation of the Japanese 9-item shared decision-making questionnaire and its association with decision conflict and patient factors in Japanese primary care. *Jma J*. 2020;3(3):208–15.
 137. Kriston L, Härter M, Scholl I. A latent variable framework for modeling dyadic measures in research on shared decision-making. *Z Evid Fortbild Qual Gesundhw*. 2012;106(4):253–63.
 138. Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item shared decision making questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Educ Couns*. 2010;80(1):94–9.
 139. Rencz F, Tamási B, Brodsky V, Gulácsi L, Weszl M, Péntek M. Validity and reliability of the 9-item Shared Decision Making Questionnaire (SDM-Q-9) in a national survey in Hungary. *Eur J Health Econ*. 2019;20(Suppl 1):43–55.
 140. Aragones A, Schaefer EW, Stevens D, Gourevitch MN, Glasgow RE, Shah NR. Validation of the Spanish translation of the patient assessment of chronic illness care (PACIC) survey. *Prev Chronic Dis*. 2008;5(4):A113.
 141. Aung E, Ostini R, Dower J, Donald M, Coll JR, Williams GM, et al. Patient assessment of chronic illness care (PACIC) in type 2 diabetes: a longitudinal study. *Eval Health Prof*. 2016;39(2):185–203.
 142. Cramm JM, Nieboer AP. Factorial validation of the Patient Assessment of Chronic Illness Care (PACIC) and PACIC short version (PACIC-S) among cardiovascular disease patients in the Netherlands. *Health Qual Life Outcomes*. 2012;10:104.
 143. Gugiu PC, Coryn C, Clark R, Kuehn A. Development and evaluation of the short version of the Patient Assessment of Chronic Illness Care instrument. *Chron Illn*. 2009;5(4):268–76.
 144. Iglesias K, Burnand B, Peytremann-Bridevaux I. PACIC instrument: disentangling dimensions using published validation models. *Int J Qual Health Care*. 2014;26(3):250–60.
 145. Kim SH, Lee BG, Choe YH. Psychometric evaluation of patient assessment of chronic illness care among Korean cancer survivors. *PLoS ONE*. 2021;16(8):e0256119.
 146. Koley M, Saha S, Ghosh S, Nag G, Kundu M, Mondal R, et al. Patient-assessed chronic illness care (PACIC) scenario in an Indian homeopathic hospital. *J Tradit Complement Med*. 2016;6(1):72–7.
 147. Krucien N, Le Vaillant M, Pelletier-Fleury N. Adaptation and validation of the patient assessment of chronic illness care in the French context. *BMC Health Serv Res*. 2014;14:269.
 148. Maingal HT, Sokolowski I, Vedsted P. Adaptation, data quality and confirmatory factor analysis of the Danish version of the PACIC questionnaire. *Eur J Public Health*. 2010;22(1):31–6.
 149. Simonsen N, Koponen AM, Suominen S. Patients' assessment of chronic illness care: a validation study among patients with type 2 diabetes in Finland. *BMC Health Serv Res*. 2018;18(1):412.
 150. Cramm JM, Nieboer AP. Development and validation of the older patient assessment of chronic illness care (O-PACIC) scale after hospitalization. *Soc Indic Res*. 2014;116:959–69.
 151. Abdul-Razak S, Ramli AS, Badlishah-Sham SF, Haniff J. Validity and reliability of the patient assessment on chronic illness care (PACIC) questionnaire: the Malay version. *BMC Fam Pract*. 2018;19(1):119.
 152. Azam AF, Lai P, Abdullah A, Haidi Hanafi NS. Validity and reliability of the Patient Assessment on Chronic Illness Care (PACIC) questionnaire among patients with type 2 diabetes mellitus in Malaysia: English version. *Malays Fam Phys*. 2020;15(2):10–8.
 153. Fan J, McCoy RG, Ziegenfuss JY, Smith SA, Borah BJ, Deming JR, et al. Evaluating the structure of the Patient Assessment of Chronic Illness Care (PACIC) survey from the patient's perspective. *Ann Behav Med*. 2015;49(1):104–11.
 154. Gensichen J, Serras A, Paulitsch MA, Rosemann T, König J, Gerlach FM, et al. The Patient Assessment of Chronic Illness Care questionnaire: evaluation in patients with mental disorders in primary care. *Community Ment Health J*. 2011;47(4):447–53.
 155. Gibbons CJ, Small N, Rick J, Burt J, Hann M, Bower P. The Patient Assessment of Chronic Illness Care produces measurements along a single dimension: results from a Mokken analysis. *Health Qual Life Outcomes*. 2017;15(1):61.
 156. Noël PH, Jones S, Parchman ML. Patient experience in an era of primary care transformation: revisiting the PACIC. *Eur J Pers Cent Healthc*. 2016;4(3):528–40.
 157. Rick J, Rowe K, Hann M, Sibbald B, Reeves D, Roland M, et al. Psychometric properties of the Patient Assessment Of Chronic Illness Care measure: acceptability, reliability and validity in United Kingdom patients with long-term conditions. *BMC Health Serv Res*. 2012;12:293.
 158. Taggart J, Chan B, Jayasinghe UW, Christl B, Proudfoot J, Crookes P, et al. Patients Assessment of Chronic Illness Care (PACIC) in two Australian studies: structure and utility. *J Eval Clin Pract*. 2011;17(2):215–21.

159. Tusek-Bunc K, Petek-Ster M, Ster B, Petek D, Kersnik J. Validation of the Slovenian version of patient assessment of chronic illness care (PACIC) in patients with coronary heart disease. *Coll Antropol*. 2014;38(2):437–44.
160. Wensing M, van Lieshout J, Jung HP, Hermesen J, Rosemann T. The Patients Assessment Chronic Illness Care (PACIC) questionnaire in the Netherlands: a validation study in rural general practice. *BMC Health Serv Res*. 2008;8:182.
161. Ali S, Chessex C, Bassett-Gunter R, Grace SL. Patient satisfaction with cardiac rehabilitation: association with utilization, functional capacity, and heart-health behaviors. *Pat Prefer Adher*. 2017;11:821–30.
162. Lambert S, McCusker J, Belzile E, Yaffe M, Ihejirika C, Richardson J, et al. Using confirmatory factor analysis and Rasch analysis to examine the dimensionality of the Patient Assessment of Care for Chronic Illness Care (PACIC). *Qual Life Res*. 2021;30(5):1503–12.
163. Iglesias K, De Geest S, Berben L, Dobbels F, Denhaerynck K, Russell LC, et al. Validation of the patient assessment of chronic illness care (PACIC) short form scale in heart transplant recipients: the international cross-sectional bright study. *BMC Health Serv Res*. 2020;20(1):160.
164. Goetz K, Freund T, Gensichen J, Miksch A, Szecsenyi J, Steinhäuser J. Adaptation and psychometric properties of the PACIC short form. *Am J Manag Care*. 2012;18(2):e55–60.
165. Vennedey V, Dust G, Schippel N, Shukri A, Strupp J, Rietz C, et al. Patient-centered care during the last year of life: adaptation and validation of the German PACIC short form for bereaved persons as proxies (PACIC-S9-proxy). *BMC Palliat Care*. 2020;19(1):177.
166. Alharbi N, Alsukbi N, Alotabi F, Alotabi M, Alharbi N, de Lusgnian S, et al. Translation into Arabic and validation of the patient assessment of care for chronic conditions questionnaire for diabetes. *East Mediterr Health J*. 2021;27(2):142–50.
167. Malliarou M, Bakola E, Nikolentzos A, Sarafis P. Reliability and validity of the Greek translation of the patient assessment of chronic illness care + (PACIC-PLUS GR) survey. *BMC Fam Pract*. 2020;21(1):122.
168. Zeugfang D, Wisetborisut A, Angkurawaranon C, Aramrattana A, Wensing M, Szecsenyi J, et al. Translation and validation of the PACIC+ questionnaire: the Thai version. *BMC Fam Pract*. 2018;19(1):123.
169. Drewes HW, de Jong-van Til JT, Struijs JN, Baan CA, Tekle FB, Meijboom BR, et al. Measuring chronic care management experience of patients with diabetes: PACIC and PACIC+ validation. *Int J Integr Care*. 2012;12:e194.
170. Rosemann T, Laux G, Droeemeyer S, Gensichen J, Szecsenyi J. Evaluation of a culturally adapted German version of the Patient Assessment of Chronic Illness Care (PACIC 5A) questionnaire in a sample of osteoarthritis patients. *J Eval Clin Pract*. 2007;13(5):806–13.
171. Schwenke M, Welzel FD, Luck-Sikorski C, Pabst A, Kersting A, Blüher M, et al. Psychometric properties of the Patient Assessment of Chronic Illness Care measure (PACIC-5A) among patients with obesity. *BMC Health Serv Res*. 2019;19(1):61.
172. Acosta AM, Lima M, Marques GQ, Levandovski PF, Weber LAF. Brazilian version of the Care Transitions Measure: translation and validation. *Int Nurs Rev*. 2017;64(3):379–87.
173. Anatchkova MD, Barysaukas CM, Kinney RL, Kiefe CI, Ash AS, Lombardini L, et al. Psychometric evaluation of the Care Transition Measure in TRACE-CORE: do we need a better measure? *J Am Heart Assoc*. 2014;3(3):e001053.
174. Flink M, Tessma M, Cvancarova Småstuen M, Lindblad M, Coleman EA, Ekstedt M. Measuring care transitions in Sweden: validation of the care transitions measure. *Int J Qual Health Care*. 2018;30(4):291–7.
175. Parry C, Mahoney E, Chalmers SA, Coleman EA. Assessing the quality of transitional care: further applications of the care transitions measure. *Med Care*. 2008;46:317–22.
176. Bakshi AB, Wee SL, Tay C, Wong LM, Leong IY, Merchant RA, et al. Validation of the care transition measure in multi-ethnic South-East Asia in Singapore. *BMC Health Serv Res*. 2012;12:256.
177. Cao X, Chen L, Diao Y, Tian L, Liu W, Jiang X. Validity and reliability of the Chinese version of the care transition measure. *PLoS ONE*. 2015;10(5):e0127403.
178. Shadmi E, Zisberg A, Coleman EA. Translation and validation of the care transition measure into Hebrew and Arabic. *Int J Qual Health Care*. 2009;21(2):97–102.
179. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. *MedCare*. 2005;43:246–55.
180. Hwang JJ, Chung JH, Kim HK. Psychometric properties of transitional care instruments and their relationships with health literacy: Brief PREPARED and Care Transitions Measure. *Int J Qual Health Care*. 2019;31(10):774–80.
181. Yoshimura M, Sato M, Sumi N. Validity and reliability of the Japanese version of the Care Transitions Measure. *Int J Health Plann Manage*. 2017;33(2):380–90.
182. McLeod J, Stolee P, Walker J, Heckman G. Measuring care transition quality for older patients with musculoskeletal disorders. *Musculoskeletal Care*. 2013;12(1):13–21.
183. Aarons GA, Ehrhart MG, Torres EM, Finn NK, Roesch SC. Validation of the Implementation Leadership Scale (ILS) in Substance Use Disorder Treatment Organizations. *J Subst Abuse Treat*. 2016;68:31–5.
184. Hu J, Gifford W, Ruan H, Harrison D, Li Q, Ehrhart MG, et al. Translation and linguistic validation of the implementation leadership scale in Chinese nursing context. *J Nurs Manag*. 2019;27(5):1030–8.
185. Hu J, Gifford W, Ruan H, Harrison D, Li Q, Ehrhart MG, et al. Validating the Implementation Leadership Scale in Chinese nursing context: a cross-sectional study. *Nurs Open*. 2021;8(6):3420–9.
186. Mandrou E, Tsounis A, Sarafis P. Validity and reliability of the Greek version of implementation leadership scale (ILS). *BMC Psychol*. 2020;8(1):49.
187. Shuman CJ, Ehrhart MG, Torres EM, Veliz P, Kath LM, VanAntwerp K, et al. EBP implementation leadership of frontline nurse managers: validation of the Implementation Leadership Scale in acute care. *Worldviews Evid Based Nurs*. 2020;17(1):82–91.
188. Torres EM, Ehrhart MG, Beidas RS, Farahnak LR, Finn NK, Aarons GA. Validation of the Implementation Leadership Scale (ILS) with supervisors' self-ratings. *Community Ment Health J*. 2018;54(1):49–53.
189. Kobayashi Y, Watanabe K, Otsuka Y, Eguchi H, Kawakami N, Imamura K, et al. Servant leadership in Japan: a validation study of the Japanese version of the Servant Leadership Survey (SLS-J). *Front Psychol*. 2020;11:1711.
190. Rodríguez-Carvajal R, de Rivas S, Herrero M, Moreno-Jiménez B, van Dierendonck D. Leading people positively: cross-cultural validation of the Servant Leadership Survey (SLS). *Span J Psychol*. 2014;17:E63.
191. Hernandez R, Roll SC, Jin H, Schneider S, Pyatak EA. Validation of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) adapted for the whole day repeated measures context. *Ergonomics*. 2021;65(7):960–75.
192. Tubbs-Cooley HL, Mara CA, Carle AC, Gurses AP. The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. *Intensive Crit Care Nurs*. 2018;46:64–9.
193. Frilund M, Fagerström L. Validity and reliability testing of the Oulu patient classification: instrument within primary health care for the older people. *Int J Older People Nurs*. 2009;4(4):280–7.
194. Rauhalu A, Fagerström L. Determining optimal nursing intensity: the RAFAELA method. *J Adv Nurs*. 2004;45(4):351–9.
195. Junttila JK, Koivu A, Fagerström L, Haatainen K, Nykänen P. Hospital mortality and optimality of nursing workload: a study on the predictive validity of the RAFAELA Nursing Intensity and Staffing system. *Int J Nurs Stud*. 2016;60:46–53.
196. Fagerström L, Rainio AK, Rauhalu A, Nojonen K. Validation of a new method for patient classification, the Oulu patient classification. *J Adv Nurs*. 2000;31(2):481–90.
197. Andersen MH, Lønning K, Bjørnelv GM, Fagerström L. Nursing intensity and costs of nurse staffing demonstrated by the RAFAELA system: liver vs. kidney transplant recipients. *J Nurs Manag*. 2016;24(6):798–805.
198. Flo J, Landmark B, Hatlevik OE, Fagerström L. Using a new interrater reliability method to test the modified Oulu Patient Classification instrument in home health care. *Nurs Open*. 2018;5(2):167–75.
199. Borsboom D, Mellenbergh GJ, Van Heerden J. The concept of validity. *Psychol Rev*. 2004;111(4):1061.
200. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge University Press; 2011.
201. Wang Y, Yin M, Zhu S, Chen X, Zhou H, Qian W. Patient-reported outcome measures used in patients undergoing total knee arthroplasty: a COSMIN systematic review. *Bone Jt Res*. 2021;10(3):203–17.

202. Rezaei M, Kolne K, Bui S, Lindsay S. Measures of workplace inclusion: a systematic review using the COSMIN methodology. *J Occup Rehabil*. 2020;30(3):420–54.
203. Hoglund LT, Scalzitti DA, Bolgla LA, Jayaseelan DJ, Wainwright SF. Patient-reported outcome measures for adults and adolescents with patellofemoral pain: a systematic review of content validity and feasibility using the COSMIN methodology. *J Orthop Sports Phys Ther*. 2023;53(1):23–39.
204. Foley T, Vale L. A framework for understanding, designing, developing and evaluating learning health systems. *Learn Health Syst*. 2023;7(1):e10315.
205. Strating MM, Nieboer AP. Psychometric test of the Team Climate Inventory-short version investigated in Dutch quality improvement teams. *BMC Health Serv Res*. 2009;9:126.
206. Hatcher M. Survey of acute care hospitals in the United States relative to technology usage and technology transfer. *J Med Syst*. 1997;21(5):323–36.
207. Rettke H, Lehmann AI, Brauchli R, Bauer GF, Petry H, Spirig R. Capturing interprofessional collaboration between physicians and nurses in an acute care setting. A validation study of the revised German version of the Collaborative Practice Scales. *J Interprof Care*. 2020;34(2):211–7.
208. Vittadello F, Mischo-Kelling M, Wieser H, Cavada L, Lochner L, Naletto C, et al. A multiple-group measurement scale for interprofessional collaboration: adaptation and validation into Italian and German languages. *J Interprof Care*. 2018;32(3):266–73.
209. Kenaschuk C, Wilkins K, Reeves S, Zwarenstein M, Russell A. Nurse-physician relations and quality of nursing care: findings from a national survey of nurses. *Can J Nurs Res*. 2010;42(2):120–36.
210. Calderon C, Ferrando PJ, Carmona-Bayonas A, Lorenzo-Seva U, Jara C, Beato C, et al. Validation of SDM-Q-Doc Questionnaire to measure shared decision-making physician's perspective in oncology practice. *Clin Transl Oncol*. 2017;19(11):1312–9.
211. Goto Y, Miura H, Son D, Scholl I, Kriston L, Härter M, et al. Association between physicians' and patients' perspectives of shared decision making in primary care settings in Japan: the impact of environmental factors. *PLoS ONE*. 2021;16(2):e0246518.
212. Stump TE, Dexter PR, Tierney WM, Wolinsky FD. Measuring patient satisfaction with physicians among older and diseased adults in a primary care municipal outpatient setting. An examination of three instruments. *Med Care*. 1995;33(9):958–72.
213. Yoshimura M, Sato M, Sumi N. Validity and reliability of the Japanese version of the Care Transitions Measure. *Int J Health Plann Manage*. 2018;33(2):380–90.
214. Ruiz Yanzi MV, Barani MS, Franco JVA, Vazquez Peña FR, Terrasa SA, Kopitowski KS. Translation, transcultural adaptation, and validation of two questionnaires on shared decision making. *Health Expect*. 2019;22(2):193–200.
215. Broström A, Pakpour AH, Nilsen P, Hedberg B, Ulander M. Validation of CollaboRATE and SURE - two short questionnaires to measure shared decision making during CPAP initiation. *J Sleep Res*. 2019;28(5):e12808.
216. Ubbink DT, van Asbeck EV, Aarts JWM, Stubenrouch FE, Geerts PAF, Atsma F, et al. Comparison of the CollaboRATE and SDM-Q-9 questionnaires to appreciate the patient-reported level of shared decision-making. *Patient Educ Couns*. 2022;105(7):2475–9.
217. Ousseine YM, Durand MA, Bouhnik AD, Smith A, Mancini J. Multiple health literacy dimensions are associated with physicians' efforts to achieve shared decision-making. *Patient Educ Couns*. 2019;102(11):1949–56.
218. Brodney S, Fowler FJ Jr, Barry MJ, Chang Y, Sepucha K. Comparison of three measures of shared decision making: SDM process_4, CollaboRATE, and SURE scales. *Med Decis Making*. 2019;39(6):673–80.
219. Forcino RC, Bustamante N, Thompson R, Percac-Lima S, Elwyn G, Pérez-Arechaederra D, et al. Developing and pilot testing a Spanish translation of CollaboRATE for use in the United States. *PLoS ONE*. 2016;11(12):e0168538.
220. De Las Cuevas C, Mundal I, Betancort M, Lara-Cabrera ML. Assessment of shared decision-making in community mental health care: validation of the CollaboRATE. *Int J Clin Health Psychol*. 2020;20(3):262–70.
221. Forcino RC, Thygeson M, O'Malley AJ, Meinders MJ, Westert GP, Elwyn G. Measuring patient-reported shared decision-making to promote performance transparency and value-based payment: assessment of collaboRATE's group-level reliability. *J Patient Exp*. 2020;7(5):742–8.
222. Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. *J Med Internet Res*. 2014;16(1):e2.
223. Campbell J, Narayanan A, Burford B, Greco M. Validation of a multi-source feedback tool for use in general practice. *Educ Prim Care*. 2010;21(3):165–79.
224. Charalambous A. Validation and test-retest reliability of the Risser patient satisfaction scale in Cyprus. *J Nurs Manag*. 2010;18(1):61–9.
225. Charalambous A, Adamakidou T. Risser patient satisfaction scale: a validation study in Greek cancer patients. *BMC Nurs*. 2012;11:27.
226. Ventura MR, Fox RN, Corley MC, Mercurio SM. A patient satisfaction measure as a criterion to evaluate primary nursing. *Nurs Res*. 1982;31(4):226–30.
227. Biering P, Becker H, Calvin A, Grobe SJ. Casting light on the concept of patient satisfaction by studying the construct validity and the sensitivity of a questionnaire. *Int J Health Care Qual Assur Inc Leadersh Health Serv*. 2006;19(2–3):246–58.
228. Yellen E, Davis GC, Ricard R. The measurement of patient satisfaction. *J Nurs Care Qual*. 2002;16(4):23–9.
229. Kasper J, Hoffmann F, Heesen C, Köpke S, Geiger F. Completing the third person's perspective on patients' involvement in medical decision-making: approaching the full picture. *Z Evid Fortbild Qual Gesundhw*. 2012;106(4):275–83.
230. Carryer J, Budge C, Hansen C, Gibbs K. Modifying the PACIC to assess provision of chronic illness care: an exploratory study with primary health care nurses. *J Prim Health Care*. 2010;2(2):118–23.
231. Gugiu C, Coryn CL, Applegate B. Structure and measurement properties of the Patient Assessment of Chronic Illness Care instrument. *J Eval Clin Pract*. 2010;16(3):509–16.
232. Choi S-Y. The Effects of leadership on social workers' organizational commitment: focusing on ethical leadership, organizational trust, and supervisor trust. *Int Inform Instit Inform*. 2014;17(10):4771.
233. Asif M, Jameel A, Hussain A, Hwang J, Sahito N. Linking transformational leadership with nurse-assessed adverse patient outcomes and the quality of care: assessing the role of job satisfaction and structural empowerment. *Int J Environ Res Public Health*. 2019. <https://doi.org/10.3390/ijerph16132381>.
234. Finn NK, Torres EM, Ehrhart MG, Roesch SC, Aarons GA. Cross-validation of the Implementation Leadership Scale (ILS) in child welfare service organizations. *Child Maltreat*. 2016;21(3):250–5.
235. Lyon AR, Cook CR, Brown EC, Locke J, Davis C, Ehrhart M, et al. Assessing organizational implementation context in the education sector: confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. *Implement Sci*. 2018;13(1):5.
236. Stogdill R. Manual for the leadership behaviour development questionnaire: form XII Columbus: Bureau of business research. Ohio: Ohio State University; 1963.
237. Carswell CM, Lio CH, Grant R, Klein MI, Clarke D, Seales WB, et al. Hands-free administration of subjective workload scales: acceptability in a surgical training environment. *Appl Ergon*. 2010;42(1):138–45.
238. Hoonakker P, Carayon P, Gurses A, Brown R, McGuire K, Khunlertkit A, et al. Measuring workload of ICU nurses with a questionnaire survey: the NASA task load index (TLX). *IIE Trans Healthc Syst Eng*. 2011;1(2):131–43.
239. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. Validation of the National Aeronautics and Space Administration Task Load Index as a tool to evaluate the learning curve for endoscopy training. *Can J Gastroenterol Hepatol*. 2014;28(3):155–9.
240. Liljamo P, Kinnunen UM, Ohtonen P, Saranto K. Quality of nursing intensity data: inter-rater reliability of the patient classification after two decades in clinical use. *J Adv Nurs*. 2017;73(9):2248–59.
241. Liljamo P, Kinnunen UM, Saranto K. Healthcare professionals' views on the mutual consistency of the Finnish classification of Nursing Interventions and the Oulu patient classification. *Scand J Caring Sci*. 2016;30(3):477–88.

242. van Oostveen CJ, Ubbink DT, Mens MA, Pompe EA, Vermeulen H. Pre-implementation studies of a workforce planning tool for nurse staffing and human resource management in university hospitals. *J Nurs Manag.* 2016;24(2):184–91.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.