

Measurement of contextual factors in clinical practice and research: development of an expert-informed toolkit

Jennifer A. Bent¹ · Chad E. Cook · Antoine Bailliard · Joel E. Bialosky · Felicity L. Bishop · Elisa Carlino · Luana Colloca · Jorge E. Esteves · Alessandra Garcia · Amy W. McDevitt · Vitaly Napadow · David Newell · Alvisa Palese · William R. Reed · Giacomo Rossetini

Received: 3 July 2026 / Accepted: 16 September 2026
© The Author(s) 2026

Abstract

Contextual factors (CFs) can influence therapeutic encounters and clinical outcomes, yet their measurement remains fragmented, lacking a unified framework to guide tool selection. To identify, appraise, and synthesize CF measurement instruments into an expert-informed clinical toolkit and identify current measurement gaps. An international panel of 15 experts evaluated 227 instruments across five domains (patient/practitioner characteristics, relational dynamics, environment, and treatment) using a hybrid COSMIN-informed appraisal design. Instruments were screened for content validity and psychometric robustness. Twenty instruments were selected for inclusion in the toolkit. Despite identifying high-quality instruments for stable traits, our analysis revealed critical measurement gaps in dynamic, state-responsive processes. Specifically, there is an absence of validated, clinically feasible tools for capturing real-time clinician affective states, non-verbal synchrony, therapeutic touch, and moment-to-moment environmental or procedural interactions. This study provides the first expert-informed framework for assessing CFs. Future research should prioritize developing state-responsive instruments capable of capturing these dynamic elements as they unfold in real-world settings to better understand the contextual drivers of recovery.

Introduction

Contextual factors (CFs) shape how healthcare interventions are delivered, interpreted and experienced¹. Clinical outcomes do not arise from treatment in isolation, but from the interaction between its technical components and the personal, relational, organizational and cultural context of care². Although trials seek to isolate treatment-specific effects, substantial variability in outcomes remains unexplained, with recent meta-analytic evidence demonstrating up to 85% of the total treatment effect in musculoskeletal pain interventions is not attributable to the specific mechanical intervention³. Importantly, this variability may reflect natural history, regression to the mean, symptom fluctuations, co-interventions, measurement-related factors and contextual influences⁴. Yet CFs

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



are rarely measured and have often been treated as background noise. Recent systematic review evidence suggests that contextual effects may contribute substantially to short-term treatment outcomes, accounting for up to 39% of observed effectiveness in some settings⁵. When left unmeasured, an important component of the clinical ecosystem may therefore remain systematically unaccounted for.

The challenge with CFs is being able to accurately measure and account for these interactions during a traditional patient-clinician encounter¹. While clinical care encompasses a variety of broader ecosystems, the focus of this toolkit is on the physical, psychological, social, and environmental elements surrounding a therapeutic encounter that can influence how a patient experiences care and how effective a treatment becomes^{1,2}. Although debate persists regarding the magnitude of contextual influences, their capacity to moderate clinical outcomes represents a critical priority across musculoskeletal research³. Because CFs represent dynamic, multi-layered phenomena, they remain notoriously complex and difficult to comprehensively measure^{3,4}.

In clinical practice and research, the inability to appropriately measure the constructs associated with CFs creates fundamental gaps in understanding their role in recovery, limit our ability to intentionally modify them as a treatment enhancer, and obscures how their influence evolves over time. This underscores the need for measurement tools that are dependable, valid, and responsive, capable of capturing CFs in a way that is both clinically meaningful and translational across settings and therapeutic encounters. The objectives of this project were to: 1) populate a comprehensive catalogue of CF measurement tools used in research and clinical practice, 2) synthesize these instruments into a CFs toolkit following targeted psychometric assessment, and 3) identify critical measurement gaps across current CFs instrument domains. The study involved an international group of experts in CFs and was performed under the guidance of a National Institutes of Health U24 initiative designed to better understand CFs and their influence on clinical outcomes.

Methods

Study design

This study employed a hybrid expert-informed instrument identification and COSMIN-informed psychometric evaluation design, integrating structured expert elicitation, targeted domain-specific searches, and a standardized algorithm to identify, appraise, and synthesize measurement instruments for CFs^{6,7}.

Study rationale

This hybrid design was selected because it provides a rigorous, transparent, and efficient way to identify and evaluate measurement instruments for CFs in areas where the evidence base is fragmented, terminology is inconsistent, and expert knowledge is essential to ensure conceptual validity⁸.

Conceptual framework

CFs were framed to a multilevel taxonomy (Fig. 1) that was derived from previous established work by Cook and colleagues⁹ and Bishop et al.¹⁰. The study framework also included the addition of a practitioner context to mirror patient context. The final framework was organized into five primary measurement domains (patient characteristics, practitioner characteristics, relational dynamics, environmental context and treatment context), that collectively evaluated both stable (enduring, trait-level) and dynamic (momentary, state-level) influences of CFs. Stable characteristics refer to fixed attributes such as personality traits or chronic beliefs, whereas dynamic characteristics refer to fluctuating, encounter-specific states such as acute or real-time expectations.

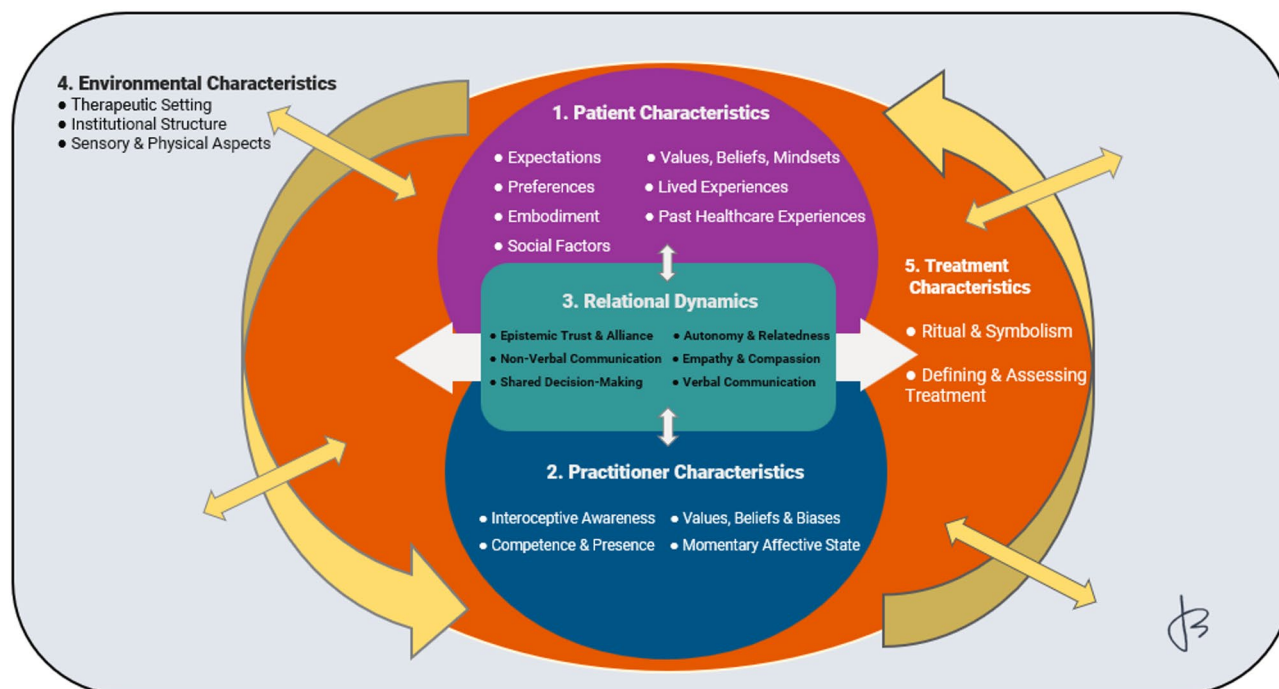


Fig. 1 The ecology of effectiveness: a multilevel taxonomy of the clinical encounter. A conceptual schematic mapping the multidimensional contextual factors that mediate and dictate clinical outcomes. The vertical axis illustrates the interaction between internal cognitive states, anchoring Patient Characteristics against Practitioner Characteristics. The horizontal axis represents the Patient-Practitioner Dynamic, functioning as the relational and non-verbal bridge these interpersonal mechanisms are inextricably nested within the external envelope of Environmental Characteristics and Treatment Characteristics. This ecological model demonstrates that therapeutic efficacy emerges not from isolated biological mechanisms, but from the systemic interaction of these five contextual domains.

Intended outputs

The study output will include a comprehensive catalogue of CF measurement tools (a toolkit) demonstrating strong psychometric performance and practical applicability in clinical settings.

Ethical approval and consent to participate

This study was reviewed by the Duke University Health System Institutional Review Board and was determined to be exempt from human subjects' research (Protocol Pro00117707). In accordance with this determination, the requirement for informed consent was waived. All methods were conducted in strict accordance with relevant institutional policies, federal regulations and applicable guidelines for research involving human participants. The study was conducted in alignment with ethical principles of beneficence, respect for persons, and justice as outlined in the Belmont Report, ensuring that the expert-informed identification and appraisal process maintained the highest standards of research integrity.

Expert panel composition

In addition to the study coordinating team (JB, GR, and CC), expert panelists were targeted based on publication record and expertise in CF research, diversity of professional discipline, and/or prior collaboration in CF research. All members were invited to participate by email. The expert panel and study coordinators (Table 1), hereby termed "panelists" comprised of 15 individuals from four countries with interdisciplinary expertise in

Table 1 Participant characteristics.

Gender	Professional background	Scientific training & background	Location	Publications related to contextual factors
Male	Chiropractor/Director of PhD in Rehabilitation Sciences Program	Anatomical Sciences/Neurobiology	United States	34
Male	Musculoskeletal Researcher	Clinical and Experimental Sciences	United Kingdom	16
Male	Osteopath	Cognitive Science	Portugal	19
Female	Physiotherapist	Clinical Research/Physiotherapy Research	United States	5
Male	Physiotherapist/Health Services Researcher	Economics	United States	28
Female	Distinguished Professor and Director	Medicine, Neuroscience and Bioethics	United States	224
Female	Professor in Nursing Science	Nursing Science	Italy	15
Male	Professor in Occupational Therapy	Occupational Science	United States	29
Male	Professor of Physical Medicine and Rehabilitation	Pain Science and Neuroimaging	United States	25
Female	Physiotherapist	Physical Therapy	United States	5
Female	Physiotherapist	Physical Therapy Rehabilitation/ Public Health (Family's Health)	United States	10
Female	Professor of Physiology & Neuroscientist	Psychology	Italy	50
Female	Professor of Health Psychology	Psychology	United Kingdom	43
Male	Professor of Physical Therapy	Rehabilitation Science	United States	34
Male	Physiotherapist	Rehabilitation Science	Italy	41

CFs, placebo research, and social determinants of health, representing research focuses on rehabilitation science, psychology, neurobiology, cognitive science, public health, clinical and experimental sciences, medicine, neuroscience and bioethics, pain science and neuroimaging, economics, nursing science, and occupational science. Mean publications in CFs were 38 (SD = 53); median = 28.

Panelist training

Panelists completed a standardized training protocol and materials including a detailed reference guide outlining the 11 psychometric and utility domains, a recorded instructional module, and a standardized slide deck based on COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Rubric^{8,9}.

Fidelity checks

The study was conducted from August 2025 to February 2026. Upon assignment of domains, researchers received monthly “check-ins” by the study coordinating team (JB, GR, and CC) and were reminded of the milestones and study timelines.

Reflexivity and panel composition

To address potential sources of interpretative bias, the research team used an iterative, reflexive process throughout the study. This involved ongoing email exchanges, virtual meetings, and structured discussions to identify, discuss, and select candidate instruments. Although the panel included professionals from multiple disciplines and countries, many members had substantial expertise in CFs, MSK care, rehabilitation, pain, and related research (Table 1). This concentration of expertise provided substantial domain-specific knowledge but may also have influenced the constructs and instruments prioritized during the appraisal process. The study was primarily designed around CFs operating within patient–clinician therapeutic encounters, and the composition of the panel reflected this focus.

Instrument identification strategy

Initially, panelists were asked to identify their domains (e.g., patient characteristics, practitioner characteristics, relational dynamics, environmental context and treatment context) and sub-domains (e.g., expectations and preferences, values, beliefs, and mindsets, etc.) of interest for a focused of recommended CF measurement tools. A comprehensive list of all domains and subdomains is provided as a supplementary file to enhance the transparency and replicability of the identification process.

Panelists were instructed to identify both “gold-standard” instruments and emerging instruments, without restrictions on publication date or indexing status. We adopted an “expert-guided identification strategy”, which is an iterative process in which domain experts help identify, nominate, and refine the pool of potential measurement instruments^{11,12}. Experts contributed their disciplinary knowledge, familiarity with emerging or unpublished tools, and awareness of instruments that may not be well indexed in traditional databases to complement targeted literature searches and ensure that the resulting instrument list is comprehensive, conceptually aligned, and relevant to real-world practice¹³.

Inclusion and exclusion criteria

We used a backwards hierarchical strategy when identifying potential CF measurement tools (Fig. 2) that was designed to remove CF measurement tools that failed to meet a priori criteria. Step one included an evaluation of instrument content validity, using the COSMIN Content Validity Rating Form^{6,7}. The COSMIN Content Validity Form is measured from 0 to 6 scale, with 0 reflecting not relevant / not clear / not comprehensive and 6 reflecting highly relevant / very clear / fully comprehensive. Tools that scored a 4/6 or higher were moved to a second step, which included a cumulative psychometric score rating. We selected this measure as a “pass through” since it reflects the fundamental requirement of whether the tool adequately measures the construct it claims to measure^{6,7}.

The cumulative psychometric score (step two) was evaluated using the Quality Criteria for Measurement Properties¹². This method is a psychometric evaluation system that assigns ratings across multiple measurement domains, such as content validity, reliability, and responsiveness, and sums these ratings to produce a total score ranging from 0 to 42. Higher summative values reflect stronger psychometric utility. Each domain is scored on a 0–3 scale, allowing reviewers to quantify the overall methodological and psychometric strength of a measurement instrument^{6,7}. This scoring approach was widely used in early measurement-quality reviews and provides a structured way to compare instruments, although it has since been largely complemented by more detailed frameworks such as COSMIN⁷. Top scoring tools were moved to step three.

Step three involved looking at individualized areas such as clinical utility, relevance, precision, sensitivity and cross-cultural validity. Judgement was made by a sub-group of panelists and study coordinators (JB, GR, and CC) using a Multi-Criteria Decision Analysis (MCDA) framework to evaluate instruments across all individual measurement domains for their comprehensive psychometric profile^{14,15}. If instruments remained tied after the

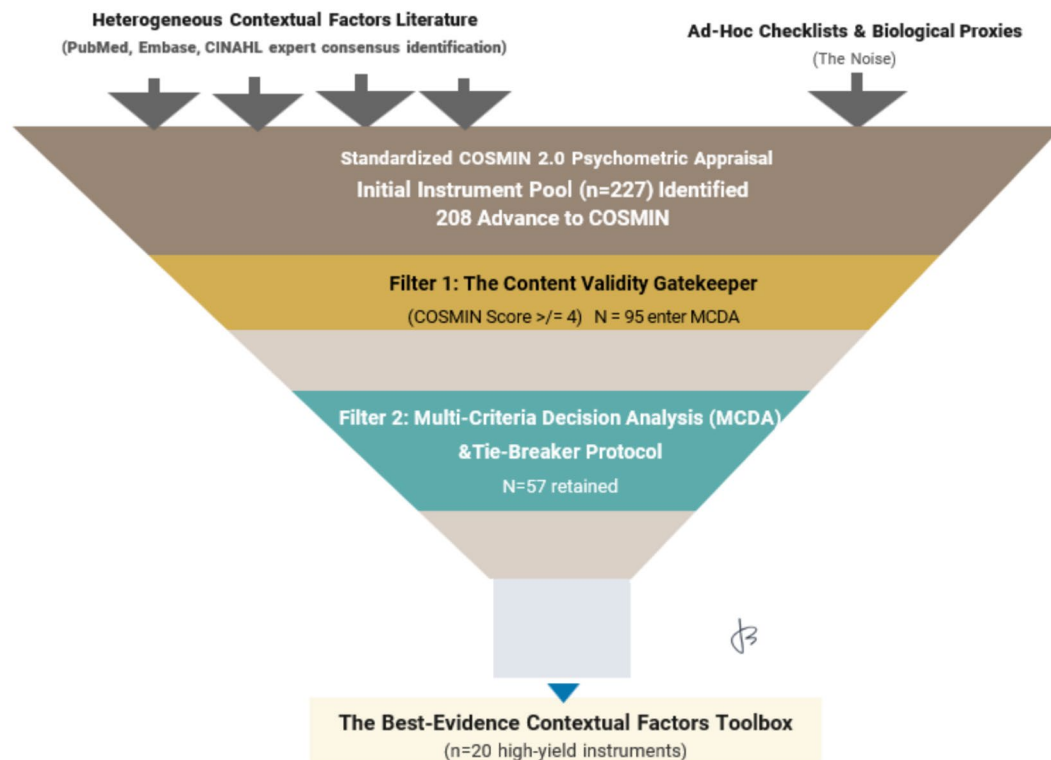


Fig. 2 Instrument screening & selection process. An initial pool of 227 instruments was evaluated using a strict Content Validity Gatekeeper (COSMIN score of ≥ 4). Instruments that passed were then processed through a Multi-Criteria Decision Analysis (MCDA); this protocol prioritized clinical utility, relevance, and precision to finalize recommended instruments. The final synthesis yielded 57 robust instruments, of which 20 were selected for primary toolkit, while all 57 are listed in Appendix A.

review, a lexicographic decision rule¹⁶ was applied. (Step Four). If ties persisted, final selection was based on practical heuristics such as open-access availability, lower levels of administrative burden, and prior panel familiarity. The final recommended tools were populated in a toolkit.

Scoring and agreement procedures

Instruments were scored by panelists using the standardized psychometric framework. Formal inter-rater reliability was not performed, as the primary aim was to enable broad domain coverage and efficient evaluation over a large and heterogeneous set of instruments.

In a small, ad-hoc instances where a tool was incidentally evaluated by more than one panelist, scoring discrepancies were managed by adopting COSMIN’s methodology’s “worst score counts” principle. Specifically, the lower score was retained, representing a conservative estimate of instrument performance and preventing overestimation of measurement quality⁶.

When multiple instruments met similar thresholds, final selection in the toolkit was determined by consensus discussion, prioritizing psychometric strength, conceptual alignment and feasibility.

Final toolkit review

Following structured filtering, the preliminary set of instruments was reviewed by the full panel to ensure domain coverage and completeness. This review resulted in the addition of one instrument based on consensus to address a gap in the measurement between empathy and communication. All instrument data, scoring decisions and

documentation were stored in a central repository and are publicly available via the study's Open Science Framework (OSF) registry: <https://osf.io/jf2sg/files/osfstorage>.

Instrument mapping

Each instrument was assigned to one of the five measurement domains based on its primary conceptual focus and the specific CF constructs it was designed to assess. This classification ensured that instruments were evaluated within conceptually coherent categories, allowing for clearer comparisons, reducing construct overlap, and strengthening the interpretability of the overall measurement framework.

Gap analysis

We evaluated potential gaps in the literature by reviewing the synthesis and mapping and evaluating where limited representation was present. To reduce the risk of oversight during our earlier searches, a targeted supplementary search was conducted to address these gaps, with newly identified tools appraised using the same COSMIN-based criteria^{6,7}.

Results

Instrument selection overview

Panelists identified 227 potential instruments. This initial pool was reduced using our hierarchical screening stages, leaving 57 instruments comprehensive individual assessment. From this pool, twenty primary instruments were selected for integration into the clinical toolkit to ensure optimal coverage across the five CF domains. The recommended tools represent both a domain and sub-domain, include scoring metrics, and include recommendations for clinical application (Table 2). For transparency and replicability, full psychometric evaluations for the initially selected 57 instruments, including highly evaluated trait measures such as the Life Orientation Test¹⁷ and the Generalized Self Efficacy Scale¹⁸ are provided in Appendix A. Inclusion and exclusion criteria are details in Appendix B.

Domain 1: Patient characteristics

This domain synthesizes the pre-existing attributes, internal baselines and cognitive mindsets that patients bring to the clinical encounter. The appraisal identified five distinct instruments required to map this domain. The Expectations for Treatment Scale¹³ provides high-fidelity baseline assessments of patients' embodied cognitions prior to an intervention. To capture stable affective and psychological traits, the Hospital Anxiety and Depression Scale (HADS)²⁰ was selected for its robust phenotypical triaging. Crucially, to evaluate the patient's embodied lived experience, the Central Sensitization Inventory (CSI)²⁰ was integrated to identify highly sensitized nervous systems that may require a shift in mechanical loading. Finally, to evaluate the historic and systemic context that brings a patient into the room, the HCAHPS survey¹⁸ and the Discrimination in Medical Settings (DMS) Scale³⁸ captures baseline comfort, medical mistrust and perceived systemic bias.

Domain 2: Practitioner characteristics

This domain accounts for the internal attributes, cognitive frameworks and momentary affective states of the clinician. The appraisal identified four primary instruments within this domain that dictate how a clinician shapes and delivers this interpersonal context.

Table 2 The contextual factors toolkit: recommended instruments for the assessment of the clinical ecosystem.

Domain	Instrument	Score ^a	Clinical application
1. Patient characteristics			
<i>This domain encompasses the lived experience, beliefs and pre-existing attributes the patient brings to the encounter</i>			
1A: Expectations & preferences <i>Pre-existing cognitive and embodied expectations regarding treatment, outcomes, and prognosis. Distinct from treatment-specific encounter-elicited expectancy</i>	Expectations for treatment scale ¹⁹	32/42	<i>Deploy at baseline:</i> Captures the cognitive and embodied expectations a patient brings into the encounter prior to intervention
1B: Values, beliefs & mindsets <i>Spirituality, anxiety, catastrophizing, cultural beliefs</i>	HADS (Hospital anxiety and depression scale) ²⁰	37/42	<i>Triage only:</i> Highly stable phenotypic measure. Excellent for identifying generalized distress, but resistant to short-term change
1C: Lived experience & embodiment <i>A focus on the patient's embodied experience influencing outcomes</i>	CSI (Central sensitization inventory) ²¹	36/42	<i>Identify threat:</i> Crucial for identifying patients whose nervous systems are highly sensitized, requiring a shift away from aggressive mechanical loading
1D: Past healthcare experiences <i>Positive or negative clinical history and comfort/trust prior to current healthcare session</i>	HCAHPS (Hospital consumer assessment of healthcare providers and systems) ²²	35/42	<i>Systematic audit:</i> Captures historical systemic interactions and baseline comfort with healthcare environments. Useful for triaging underlying medical mistrust before treatment begins
1E: Sociodemographic factors & social Factors <i>Social determinants of health, access barriers, discrimination experiences, and community engagement</i>	Discrimination in medical settings (DMS) Scale ²³	33/42	<i>Contextual triage:</i> Identifies stable social determinants and perceived systemic bias. Highly resistant to short-term intervention but critical for establishing baseline cultural safety
2. Practitioner characteristics			
<i>Accounting for the clinician's embodied cognition, competence and momentary affective state</i>			
2A: Interoceptive Awareness <i>The clinician's own lived experience, burnout, and embodied cognition</i>	MAIA-2 (Interoceptive awareness – self regulation) ¹⁸	38/42	<i>Self-audit:</i> Measures the clinician's own embodied cognition and ability to self-regulate during complex or stressful clinical encounters
2B: Values, beliefs & biases <i>Clinical preferences, implicit bias, stigma and workplace challenges</i>	HC-PAIRS (Providers' pain & impairment) ²⁴	39/42	<i>Identify bias [Highest scoring]:</i> Exposes rigid clinical preferences and phenotypic biases that may subconsciously alter treatment delivery
2C: Competence & clinical presence <i>Integration of professional experience, decision-making process, and use of evidence-based practices (EBP) that fosters epistemic trust, distinct from technical proficiency</i>	Fresno test ²⁵	25/42	<i>Knowledge audit:</i> Evaluates stable EBP competence. Trade-off: The lower overall score (25/42) highlights a persistent measurement void in capturing real-time, dynamic clinical reasoning during an encounter that fosters patient trust
2D: Momentary Affective State <i>Situational awareness and emotional state during the encounter</i>	PANAS-SF (Positive and negative affect) ²⁶	39/42	<i>Encounter tracking:</i> A rare, highly responsive tool capable of capturing a clinician's acute emotional state immediately surrounding a specific session
3. Relational dynamics			
<i>Capturing the relational "space between" including alliance, trust, communication (verbal/non-verbal)</i>			
3A: Epistemic trust & alliance <i>Fostering robust therapeutic alliances, cultural humility and interpersonal trust, relational safety</i>	Wake forest scale (Competence subscale) ²⁷	39/42	<i>The defensive playbook:</i> The foundational metric for establishing baseline interpersonal trust and cultural humility
3B: Autonomy & relatedness support <i>Autonomy, competence, relatedness, situational awareness for interpersonal dynamics</i>	Group environment questionnaire (GEQ) ²⁸	37/42	<i>Dynamic tracking:</i> Captures situational awareness and the sense of relatedness/autonomy within the therapeutic environment. Highly sensitive to encounter-specific relational shifts in the dyad
3C: Non-Verbal Communication <i>Touch, posture, and other non-verbal cues that shape rapport and relational meaning</i>	NIS-S (Nonverbal immediacy scale) ²⁹	32/42	<i>Relational tracking:</i> Quantifies the often-ignored relational impact of physical touch, eye contact, and physical proximity in the clinical dyad

Table 2 (continued)

Domain	Instrument	Score ^a	Clinical application
3D: Empathy & compassion <i>The bidirectional exchange and perception of emotional understanding and care</i>	Compassionate care assessment tool (CCAT) ³⁰ Consultational and relational empathy (CARE) measure ³¹	35/42 30/42	<i>Compassion enactment (CCAT)</i> : Quantifies the situational, patient-perceived enactment of compassionate behaviors rather than the clinician's internal disposition. Critically surmounts the restricted score variability of legacy trait measures (e.g., Jefferson Scale) by capturing the actionable, moment-to-moment therapeutic bridge <i>Empathy perception (CARE)</i> : Evaluates the vital patient-reported experience of clinical empathy. Serving as a primary partner to the CCAT, this tool captures the subjective relational safety felt by the patient, which is essential for identifying mechanisms of contextual modulation
3E: Shared decision making <i>Building connections through collaborative approaches</i>	Decisional conflict scale (DCS) ³²	34/42	<i>Actionable tracking</i> : Responsive tool that quantifies acute decisional uncertainty, signaling the clinician to pause physical intervention and pivot to relational or educational strategies
3F: Verbal Communication <i>Clarity, tone, and content of spoken communication</i>	Health care communication questionnaire (HCCQ-Comm) ³³	21/42	<i>Critical measurement void</i> : Evaluates patient perception of communication. Trade-off: The very low methodological score reveals a critical lack of rigorous, highly validated instruments to measure the actual nuance of clinical dialogue
4. Environmental context <i>The structural and sensory environment of the healthcare system</i>			
4A: Therapeutic setting & institutional structure <i>Includes reimbursement, decision-making hierarchies, trainee involvement, telehealth, and organizational culture</i>	Perceived organizational support (POS)- 8 Item ³⁴	36/42	<i>Systematic audit</i> : Evaluates the overarching institutional climate and clinician support rather than the immediate clinical room. Highly stable; best suited for administrative utility and predicting burnout
4B: Sensory and physical aspects <i>Lighting, sounds, temperature, and posture/furniture in the therapeutic environment</i>	CBE indoor environmental quality (IEQ) occupant Survey ³⁵	37/42	<i>Baseline audit</i> : Robust structural evaluation of spatial layout, privacy, and clinical lighting. Highly stable; use for structural clinic design rather than session-by-session tracking
5. Treatment context <i>The symbolic and ritualistic elements of the intervention itself, distinct from its specific bio-physiological mechanism</i>			
5A: Ritual & symbolism <i>Rituality of treatment actions, sensory intensity, invasive vs. non-invasive symbolism</i>	TCS (Treatment credibility scale) ³⁶	33/42	<i>Assess cognitive buy-in</i> : effectively separates the perceived logical coherence and symbolism of an intervention from its actual biological mechanism
5B: Defining and assessing treatment <i>Elements such as touch, rituality, and mechanistic understanding when supported by prior evidence</i>	Revised illness perception questionnaire (IPQ-R) – Treatment Control/coherence ³⁷	31/42	<i>Cognitive tracking</i> : Measures the shift in how much control and coherence a patient feels regarding their specific treatment plan as the educational intervention progresses

^a Total Score reflects the combined methodological quality across all COSMIN 2.0 parameters.

Superscript numbers indicate the primary validation or anchor studies corresponding to the manuscript's master reference list.

The Fresno Test remains the most validated measure of evidence-based practice (EBP) competence across healthcare professions²⁵. However, because it evaluates stable knowledge rather than real-time, dynamic clinical reasoning during an encounter, a measurement void persists in capturing how competence is actively signaled to the patient mid-encounter to foster trust. Similarly, the Health Care Provider's Pain and Impairment Relationship Scale (HC-PAIRS)³⁹ remains the primary instrument for identifying rigid clinical preferences and historical biases regarding pain and disability that may subconsciously alter treatment delivery. To address the clinician's own embodied cognition and self-regulation, the MAIA-2 (Self-Regulation Subscale)²⁴ was identified as a superior tool for monitoring interoceptive awareness during complex encounters.

This gap is consequential when a clinician experiences real acute, situational burnout, and/or emotional depletion can immediately neutralize a clinician's stable empathy or communication traits. An acutely depleted clinician fundamentally alters the immediate interpersonal input of the session, changing how a patient perceives

relational safety. The PANAS-SF²⁶ was therefore integrated into the toolkit as a rare, responsive proxy for capturing a clinician's acute emotional state immediately surrounding a specific session. While tools like the PANAS-SF capture broad affect, the field still lacks validated instruments capable of measuring a clinician's real-time cognitive load or situational awareness mid-encounter. Even with the best available tools, the absence of a comprehensive measure that answers the practical question "What is the clinician's state right now?", represents a critical measurement deficit.

Domain 3: The patient–practitioner dynamic

This domain captures the relational space between the horizontal axis of our taxonomy encompassing the mutual trust, alignment, and verbal/non-verbal communication generated within the clinical dyad. The appraisal identified seven instruments required to map this domain.

Because this domain is transactional, it introduces a methodological tension, whether interpersonal dynamics are best captured by practitioner self-report, patient perception or objective coding.

The Wake Forest Physician Trust Scale³⁹, and the Group Environment Questionnaire (GEQ)²⁸ was identified as the highest performing instruments for establishing relational safety and situational alignment within the encounter. To evaluate clinical empathy, legacy self-report measures like Jefferson Scale of Empathy (JSE-HP)²⁷ often fail to capture the patient's actual lived experience. To resolve this, the patient reported CARE measure³¹ was integrated into the toolkit to evaluate the dynamic, patient perceived enactment of empathy, while the Compassionate Care Assessment Tool (CCAT)³⁰ remains a critical adjunct to capture encounter-specific compassion.

To capture the nonverbal dimension, the Nonverbal Immediacy Scale-Self Report (NIS-S)²⁹, provides a validated method for quantifying warmth cues such as gaze, posture and facial expression. The Decisional Conflict Scale (DCS)³², demonstrates strong psychometric performance for shared decision making for when a clinician must pivot away from clinical procedures and towards relational or educational support. Finally, for verbal communication, the Health Care Communication Questionnaire (HCCQ-Comm)³⁰ was mapped to the toolkit, however it's low methodological score reveals a lack of rigorous instruments designed to measure the nuance of actual clinical dialogue.

Domain 4: Characteristics of the setting

This domain highlights the clinical significance of both macro-level institutional and micro-level sensory cues within the healthcare environment, which represents the external structure of our taxonomy. To capture the broader institutional climate, the Perceived Organizational Support (POS) 8-Item Scale³⁴ emerged as the strongest tool for evaluating organizational culture, decision-making hierarchies, and clinician support. This metric is critical for identifying structural environments predisposed to clinician burnout. By contrast, tools evaluating the immediate physical room such as the CBE-IEQ³⁵, provides a structurally robust measurement of clinical design, spatial layout, and patient privacy.

While these structural tools form the baseline toolkit, the appraisal also highlighted critical, previously under-described sensory considerations within the clinical ecosystem. Although often treated as isolated engineering considerations, acoustics function as active sensory signals that modulate autonomic stress physiology and placebo responses. The Soundscape Quality (SQ) Scale⁴⁰ offers a validated method for quantifying the pleasantness eventfulness and of auditory environments, correlating with physiological stress markers.

Furthermore, the consensus panel also distinguished between measuring pure olfactory function and clinical olfactory context. Although olfaction relies on a biological mechanism, it is uniquely positioned as a contextual driver. The trigeminal nerve relays to the spinal trigeminal nucleus and higher cortical networks. Because standard diagnostic smell tests such as UPSIT⁴¹ and Sniffin' Sticks⁴² evaluate anatomical function, they lack content validity for evaluating therapeutic environments. The Olfactory Dysfunction Outcome Rating Scale (ODOR)⁴³ provides an appropriate proxy by capturing the emotional and contextual impact of environmental scents on the

patient. While these sensory tools demonstrate strong psychometric potential and are listed in our supplementary index, they currently prioritize baseline structural design than interactive, encounter level tracking.

Domain 5: Treatment characteristics

This domain encompasses the symbolic, cognitive and ceremonial elements of the intervention itself, forming the final external layer of our taxonomy. In alignment of our framework, this domain separates the psychological meaning of an intervention from its underlying bio-physiological or mechanical pathway. The appraisal identified two instruments required to map this domain.

For evaluating cognitive “buy-in,” perceived credibility, and the symbolic logic of an intervention (the “why”), the Treatment Credibility Scale (TCS)³⁶ provides high-quality data. The TCS tracks how a patient’s perception of an intervention’s logical clarity shapes their baseline treatment expectations. Similarly, the Revised Illness Perception Questionnaire (IPQ-R) Treatment Subscale¹⁹ effectively assesses the patient’s explicit mechanistic understanding and their perceived sense of personal control over the intervention.

Gap analysis

A number of important measurement gaps emerged across the five domains, revealing substantial relational, environmental, and procedural voids. In the relational domain, a clear disparity exists between tools that assess decision-making processes (“choice”) and those that capture the nuance of moment-to-moment clinical dialogue (“chat”); although shared decision-making measures are psychometrically strong, instruments such as the Health Care Communication Questionnaire (HCCQ-Comm)³³ and the Communication Assessment Tool (CAT)⁴⁴ lack the rigor needed for routine clinical use, underscoring the need for more sensitive, state-responsive communication measures.

Parallel gaps were evident in non-verbal synchrony and therapeutic touch, where existing approaches rely on labor-intensive video-coding systems like the Roter Interaction Analysis System (RIAS)⁴⁵, limiting feasibility in real-time clinical workflows, and where tools such as the Social Touch Questionnaire (STQ)⁴⁶, retained only in Appendix A due to lower robustness, capture trait-level touch avoidance but fail to provide a viable state-responsive measure of physical interaction. Environmental measurement gaps were similarly pronounced: while legacy instruments such as HCAHPS²² demonstrate strong psychometric performance for administrative auditing, their environmental subscales overlook granular sensory variables (e.g., lighting spectrum, temperature, olfactory context) that meaningfully influence patient physiology, and even robust tools like CBE-IEQ³⁵ and ASPECT⁴⁷ remain limited by their static nature, leaving dynamic posture, proximity, and movement unmeasured.

Finally, procedural gaps were evident in the limited ability to assess the physical enactment and ritualistic qualities of treatment delivery; although instruments like the Touch & Health Scale (THS)⁴⁸ attempt to quantify interpersonal physical contact, they exhibit poor psychometric robustness, and multiple touch-related or clinical-integrity scales were excluded due to inadequate content validity. Collectively, these gaps highlight the absence of psychometrically sound, state-responsive tools capable of capturing the relational, sensory, and ritual dimensions of care that shape the therapeutic encounter.

Discussion

Objectives

The goals of this project were to: 1) populate a list of CF measurement tools used in research and clinical practice, 2) after targeted psychometric assessment, synthesize these instruments into a CFs toolkit, and 3) identify notable gaps in current CFs instrument domains. We were able to identify 227 initial instruments (objective one),

which we ultimately reduced to a core group of twenty measurement tools that represent five domains (and their subdomains) of CFs (objective two). We also identified notable gaps in instrument representation, particularly for dynamic relational, environmental, and treatment-related processes (objective three). These findings have implications for how CFs are conceptualized, measured, and incorporated into research and clinical practice.

The importance of accurate measurement and representation

Appropriately measuring CFs is essential because these elements, such as a patient's beliefs, expectations, social environment, access to resources, and therapeutic alliance, can meaningfully shape both clinical outcomes and the trajectory of recovery^{4,9,49}. When these influences are only partially captured or measured with overly narrow tools, clinicians and researchers may overlook relevant contributors to improvement or stagnation⁵⁰. Comprehensive instruments must therefore represent the full spectrum of constructs associated with CFs, ensuring that psychological, social, cultural, environmental, and interpersonal dimensions are all accounted for⁹. Without this breadth, important contributors to recovery remain invisible, limiting the accuracy of clinical reasoning and the effectiveness of interventions. Robust measurement of CFs ultimately supports more precise interpretation of patient progress, a clearer understanding of how treatment effects interact with the lived realities of the individuals we serve.

At the same time, an important consideration is that the present framework necessarily decomposes CFs into measurable constructs, whereas these factors are unlikely to operate independently within clinical encounters^{9,49}. Patient expectations, practitioner characteristics, relational processes, environmental cues, and treatment-related meanings may interact dynamically and reciprocally, with their relative influence changing across time and clinical situations. Accordingly, the toolkit should not be interpreted as representing isolated or simply additive determinants of outcomes, but rather as providing measurable components within a broader adaptive clinical system. A complexity science and systems-thinking perspective may therefore provide a complementary framework for examining how CFs interact across patient, practitioner, relational, treatment, and environmental levels. Future research should investigate not only individual CFs, but also their interactions and evolving configurations across episodes of care.

Recommended use of the toolkit

We designed this toolkit as a resource for clinicians and researchers seeking to identify measurement instruments aligned with specific CF domains and subdomains. Although our initial intention was to identify a small number of instruments per domain to maximize administrative simplicity, the final framework prioritized representation across subdomains rather than imposing an arbitrary numerical limit.

The toolkit should be considered a menu of measurement options rather than a standardized battery to be routinely administered in its entirety. Instrument selection should be guided by the clinical or research question, the contextual domain of interest, the temporal characteristics of the construct being assessed, and the feasibility of measurement within the intended setting. Thus, some instruments may be more appropriate for research, organizational evaluation, or baseline characterization, whereas others may lend themselves more readily to use during ongoing clinical care.

Researchers may use these tools to quantify CFs, examine their associations with outcomes, investigate potential mechanisms, or evaluate how contextual processes change over time. Clinicians, however, may use the toolkit more selectively to support context-informed clinical assessment. For example, treatment expectations or central sensitization inventory may be assessed at baseline when they appear particularly relevant to how a patient approaches care^{19–21}; the Decisional Conflict Scale may be considered when uncertainty regarding treatment choices emerges³²; and measures of perceived empathy, compassion, communication, or treatment credibility may be useful when relational difficulties, disengagement, or doubts regarding the proposed intervention become apparent^{30,31,33,36,51}. In this way, measurement can complement clinical reasoning and patient dialogue by making

otherwise implicit contextual influences more visible and potentially actionable. The purpose is not to quantify every CF at every encounter, but to select measurement when it addresses a specific clinical question.

This selective approach is also consistent with the systemic integration of whole-person health and self-management support in musculoskeletal care^{52,53}. For example, quantifying a patient's expectations, treatment preferences, perceived barriers, or previous healthcare experiences may help different members of a care team develop a more shared understanding of how the patient is experiencing care. Similarly, tracking measures of trust, communication, or treatment credibility does not just generate passive quantitative data, it can help identify areas that warrant further dialogue and investigation. Rather than questioning if this approach improves outcomes, the next step necessary is to evaluate how using these tools in routine workflows actively drives coordination and supports the systematic delivery of whole-person health and self-management support^{52,53}.

Methodological gaps

Contextual factors have long been treated as background “noise” rather than meaningful clinical signals, and our analysis highlights a persistent imbalance in how they are measured: stable patient traits are well captured by existing tools, while dynamic clinician behaviors, affective states, and moment-to-moment sensory features remain largely unmeasured¹. This emphasis on retrospective, easily standardized constructs overlooks evidence that many contextual influences operate within rapid, situational timeframes, meaning that instruments unable to detect these shifts may limit understanding of how contextual processes operate over time and reduce clinicians' ability to intentionally use contextual modulation in care. The sensory gap illustrates this challenge. Although static environmental features are relatively well documented, the patient's lived sensory experience remains less well captured, even though clinical environments function as active signals that can either support or undermine otherwise evidence-based interventions, helping explain variability in outcomes^{54,55}. Similarly, while symbolic meaning and expectancy are well studied, the ritual enactment of treatment (the “how” of delivery) remains poorly operationalized; existing measures capture belief and credibility but rarely isolate ritual from underlying biological mechanisms, and several tools assessing touch or procedural integrity lack adequate content validity⁵⁶. Without robust instruments that measure context as it is lived and ritual as it is enacted, the field will continue to struggle to disentangle specific treatment effects from contextual processes that may contribute to recovery.

Clinical and research implications

Transitioning CFs from abstract concepts to meaningful clinical measures requires instruments that are responsive, brief and conducive for real-time workflows. Based on the measurement gaps identified in our appraisal, research is needed for patient-reported assessments of non-verbal immediacy and therapeutic touch, sensory environmental cues, and ritualistic dimensions of treatment enactment. In addition, although the present toolkit provides a structured framework for CFs assessment, its feasibility, acceptability, implementation burden, and influence on clinical decision-making have not yet been prospectively evaluated. Future studies should therefore examine the practical integration of the toolkit in both clinical and research settings, including its usability, uptake by clinicians, and ability to inform person-centered care. Addressing these gaps is essential to understanding the true role of CFs on outcomes and recovery.

Study limitations

This study has several methodological limitations. Because expert reviewers were divided into domain-specific subgroups, not all instruments underwent dual independent appraisal, preventing calculation of formal inter-rater reliability and introducing the possibility of subgroup bias. To reduce subjective rating error, the study incorporated three safeguards: overlapping instruments that showed minimal score variation across reviewers, a conservative rule that defaulted to the lower score when discrepancies occurred, and a secondary quality-assurance step

to align qualitative notes with final scores. In addition, instrument identification was guided by expert knowledge and targeted searches rather than a comprehensive systematic review, creating the possibility that relevant instruments were overlooked.

Additionally, the framework lacked a formal consensus-driven definition for patient and clinician posture at the time of instrument screening. This definitional gap was identified after instrument screening was underway and could not be incorporated prospectively into the appraisal process. Consequently, instruments specifically measuring posture as a CF are underrepresented in the current toolkit, representing an area that requires formal categorization in future iterations.

Although the panel represented multiple disciplines and countries, its composition was concentrated in CFs, MSK care, rehabilitation, pain, and related research, which may have influenced the constructs and instruments prioritized during the appraisal process. Some perspectives, including those of patients and public representatives, generalist clinicians, nurses, and professionals working in other healthcare settings, were less represented or absent. Different healthcare ecosystems may involve additional contextual influences or operationalize relational, environmental, and organizational factors differently. Accordingly, the present findings are likely to be most directly applicable to patient–clinician interactions in MSK and rehabilitation settings. Future iterations of the toolkit should incorporate broader clinical representation and lived-experience perspectives to evaluate its relevance across other healthcare contexts. Finally, when instruments demonstrate similar scores, selection prioritized responsiveness and clinical utility over maximal precision, which supports real-world applicability but may underrepresent instruments better suited to specific research applications.

Conclusions

This study delivers a structured toolkit for identifying and selecting contextual factors (CF) measures across patient, practitioner, relational, environmental and treatment-related domains. While many of these instruments possess established psychometric properties, a critical measurement gap remains. Specifically, the field lacks tools capable of capturing dynamic clinician behaviors, moment-to-moment patient experiences, and other rapidly shifting elements of the clinical encounter.

To address this, the necessary next step is implementation validation to establish a continuous translational loop. Moving this framework from the psychometric bench to the clinical bedside will allow us to evaluate the operational feasibility across diverse practice environments. By capturing how these contextual factors dynamically interact, we can more accurately quantify their contributions to patient and clinical outcomes. This bench-to-bedside and back again approach is an essential mechanism to enforce systemic delivery and research of whole person care.

Appendix A: Full CF toolbox with 57 tools with COSMIN scoring matrix

Rank	Instrument / Sub-Domain	Content validity (0-6)	Structural validity (0-3)	Internal consistency (0-3)	Reliability (0-3)	Measurement error (0-3)	Cross-cultural validity (0-3)	Criterion validity (0-6)	Hypothesis testing (0-3)	Responsiveness (0-3)	Applicability (0-3)	Usability (0-3)	Total score (0-42)
1. PATIENT CHARACTERISTICS													
1A: Expectations & Preferences: Cognitive/embody expectations regarding treatment/prognosis													
1	Expectations for Treatment Scale	6	3	3	3	0	3	6	2	0	3	3	32
2	Illness Perception Questionnaire (IPQ-R) (Brief)	6	1	1	3	1	3	3	3	4	3	3	31
3	Tex-Q	6	3	3	2	0	3	4	2	0	3	3	29
1B: Values, Beliefs, & Mindsets: Spirituality, anxiety, catastrophizing, cultural beliefs													
1	Hospital Anxiety and Depression Scale (HADS)	6	3	3	3	3	3	4	2	4	3	3	37
2	State-Trait Anxiety Inventory (STAI – State Subscale)	5	3	3	2	3	3	5	3	5	2	2	36
3	WHOQOL-SRPB	6	3	3	3	2	3	4	3	2	3	3	35
1C: Lived Experience & Embodiment: A focus on the patient's embodied experience influencing outcomes													
1	CSI (Central Sensitization Inventory)	6	2	3	3	2	3	4	3	4	3	3	36
2	Multidimensional Assessment of Interoceptive Awareness (MAIA-2)	6	3	3	3	0	3	3	2	4	3	3	33
3	Body Perception Questionnaire-Short Form (BPQ-SF)	6	3	3	3	2	2	4	3	0	3	3	32
1D: Past Healthcare Experiences: Positive or negative clinical history and comfort with current healthcare													
1	HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems)	6	3	3	3	2	3	4	3	3	2	3	35
2	PARIS-PQ Patient Questionnaire	6	3	3	2	2	3	4	3	4	2	3	35
3	Epistemic Trust, Mistrust, and Credulity Questionnaire (ETMCQ)	6	3	3	2	1	3	4	3	3	3	3	34
4	Medical Mistrust Index (MMI)	6	2	3	2	2	3	4	3	2	3	2	32
1E: Sociodemographic & Social Factors: SDOH, barriers to access, community engagement													
1	Discrimination in Medical Settings (DMS) Scale	6	3	3	2	2	3	4	3	2	3	2	33
2	Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE)	6	3	3	0	0	0	0	3	0	3	3	21
3	Accountable Health Communities Health-Related Social Needs (AHC-HRSN)	6	0	0	0	0	0	0	0	0	3	0	9
2. PRACTITIONER CHARACTERISTICS													
2A: Clinician Context: The clinician's own lived experience, burnout, and embodied cognition													
1	The MAIA-2 – Self Regulation Subscale	6	3	3	3	2	3	5	3	4	3	3	38
2	Maslach Burnout Inventory (MBI)	6	3	3	3	2	3	6	3	3	2	3	37
3	Utrecht Work Engagement Scale (UWES-9)	6	2	3	3	1	2	6	3	4	3	3	36
4	The MAIA-2 – Body Listening Subscale	6	3	2	2	2	3	4	3	4	3	3	35
2B: Values, Beliefs & Mindsets: Clinical preferences & biases, workplace challenges													
1	Health Care Providers' Pain and Impairment Relationship Scale (HC-PAIRS)	6	3	3	3	2	3	4	3	6	3	3	39
2	Pain Attitudes and Beliefs Scale for Physiotherapists (PABS_PT)	6	3	3	2	2	3	4	3	4	2	3	35
3	Physician Belief Scale (PBS)	6	2	3	3	2	2	4	3	4	2	3	34
2C: Competence & Training: Professional experience & decision-making processes, including evidence-based practices													
1	Fresno Test	6	0	3	3	0	3	0	3	3	2	2	25
2	Evidence-Based Practice Confidence (EPIC)	6	0	3	3	0	0	0	3	6	0	3	24
3	Value, Implementation, and Knowledge of Evidence-Based Practice (QUICK-EBP-VIK)	6	0	3	3	0	3	0	0	6	3	0	24
2D: Momentary State: Current emotions & situational awareness during the encounter.													
1	PANAS-SF (Positive and Negative Affect Schedule – Short Form)	6	3	3	2	2	3	5	3	6	3	3	39
2	PDR (Pupillary Distress Response)	6	3	0	3	3	3	5	3	6	0	1	33
3	Sustained Attention to Response Task (SART)	5	3	0	3	3	3	4	2	4	0	1	28
3. CHARACTERISTICS OF THE DYNAMIC BETWEEN PATIENT & PRACTITIONER													
3A: Trust & Epistemic Trust: Fostering robust therapeutic alliances, cultural humility, interpersonal trust													
1	Wake Forest Physician Trust Scale (Competence Subscale)	6	3	3	3	2	3	5	5	3	3	3	39
2	Epistemic Trust, Mistrust, and Credulity Questionnaire (ETMCQ)	6	3	3	2	1	3	4	3	3	3	3	34
3	The Cultural Humility Scale (CHS)	6	3	3	2	1	3	5	3	2	3	3	34
3B: Self-Determination Theory Variables: Autonomy, competence, relatedness, situational awareness for interpersonal dynamics													
1	Group Environment Questionnaire (GEQ)	6	3	3	3	1	2	6	3	4	3	3	37
2	Self-Regulation Questionnaire (SRQ)	6	2	3	2	2	3	6	3	4	3	3	37
3	Task and Ego Orientation in Sport Questionnaire (TEOSQ)	6	3	3	2	1	2	6	3	4	3	3	36
3C: Non-Verbal Communication: Touch as a relational tool to build trust and rapport													
1	Nonverbal Immediacy Scale (NIS-S) – Self Report	6	3	3	3	2	3	4	3	2	2	1	32
2	The Non-verbal Accommodativeness Scale	5	2	3	2	1	1	4	2	4	3	3	30
3	Social Touch Questionnaire (STQ)	6	2	3	3	1	2	3	2	1	3	3	29
3D: Empathy & Compassion: Reflective of patient & clinician characteristics rather than purely relational dynamics													
1	The Jefferson Scale of Empathy (JSE-HP – Health Professionals Version)	6	3	3	3	2	3	5	3	3	3	3	37
2	Compassionate Care Assessment Tool (CCAT)	6	3	3	3	2	2	5	3	4	2	3	35
3	The Sinclair Compassion Questionnaire (SCQ)	6	3	3	3	0	3	6	3	0	3	3	33
3E: Shared Decision Making: Building connections through collaborative approaches													
1	Decisional Conflict Scale (DCS)	6	3	3	3	0	3	3	2	5	3	3	34
2	Observing Patient Involvement (OPTION) 12 Scale	6	3	3	2	0	3	3	3	4	2	2	31
3	CollaboRATE	6	0	0	3	2	1	4	3	6	3	3	31
3F: Verbal Communication:													
1	Health Care Communication Questionnaire (HCCQ-Comm)	6	2	3	0	0	0	4	0	0	3	3	21
2	Communication Assessment Tool (CAT)	6	3	3	0	0	0	0	2	0	3	3	20
3	Provider-Patient Relationship Questionnaire (PPRQ)	4	1	3	3	0	0	0	0	0	3	3	17
4. SETTING CHARACTERISTICS													
4A: Therapeutic Setting & Institutional Structure: Includes reimbursement, decision-making hierarchies, trainee involvement, telehealth, organizational culture													
1	Perceived Organizational Support (POS) – 8 Item	6	3	3	3	1	3	5	3	3	3	3	36
2	Mini-Z 2.0 Burden Subscale	6	3	3	3	0	3	4	3	4	3	3	35
3	Safety Attitudes Questionnaire (SAQ) – Ambulatory Version	5	3	3	3	2	3	4	3	4	2	3	35
4	System Usability Scale (SUS)	4	2	3	3	1	3	4	3	4	3	3	33
5	Dundee Ready Education Environment Measure (DREEM)	6	2	3	2	1	3	4	3	3	2	3	32
4B: Sensory and Physical Aspects: Lighting, sounds, temperature, and posture/furniture in the therapeutic environment													
1	CBE Indoor Environmental Quality (IEQ) Occupant Survey	6	3	3	3	2	3	4	3	4	3	3	37
2	A Staff and Patient Environment Calibration Tool (ASPECT)	6	3	3	3	2	2	3	3	3	2	3	33
3	Building Use Studies Occupant Questionnaire (BUS)	6	2	3	3	2	3	3	2	4	2	2	32
4	Soundscape Quality (SQ) Scale	4	3	3	3	3	2	4	3	2	3	2	32
5	ODOR (Outcomes Rating)	4	3	3	3	2	1	4	3	4	2	2	31
5. TREATMENT CHARACTERISTICS													
5A: Ritual & Symbolism: The ritual of treatment actions, sensory intensity, invasive vs. non-invasive symbolism													
1	Treatment Credibility Scale (TCS)	5	2	3	2	2	2	5	3	3	3	3	33
2	Low Back Pain Treatment Beliefs Questionnaire (LBP-TBQ)	6	2	3	3	0	2	4	3	0	3	3	29
3	Credibility/Expectancy Questionnaire (CEQ)	4	2	3	3	0	2	4	3	0	3	3	27
4	Stanford Expectations of Treatment Scale (SETS)	6	3	3	2	0	0	3	3	0	3	3	26
5B: Defining and Assessing Treatment: Exploring elements like touch, ritual, and mechanistic understanding when supported by prior research													
1	Revised Illness Perception Questionnaire (IPQ-R) – Treatment Control/Coherence	6	3	3	1	0	2	2	2	4	3	3	31
2	Patient Enablement Instrument (PEI)	6	3	3	1	0	2	2	2	4	3	3	29
3	Touch & Health Scale (THS)	4	2	3	1	2	0	4	2	0	0	2	20

Instruments are ranked within each sub-domain by Total Score (T1). COSMIN domains: CV=Content Validity (0–6); SV=Structural Validity (0–3); IC=Internal Consistency (0–3); Rel=Reliability (0–3); ME=Measurement Error (0–3); CCV=Cross-Cultural Validity (0–3); CrV=Criterion Validity (0–6); HT=Hypothesis

Testing (0–3); Resp=Responsiveness (0–6); App=Application (0–3); Acc=Acceptability (0–3); Total=Sum of all domains. *Note.* Instruments are ranked 1–N within each sub-domain by descending Total Score. Unranked entries (marked –) are additional instruments identified but not formally ranked. COSMIN scoring: Content Validity 0–6; Responsiveness 0–6; Criterion Validity 0–6; all other domains 0–3. Maximum possible Total Score=42. Score color coding: green=score $\geq$ 5; red=score=0; yellow=intermediate scores. Abbreviations: CV, content validity; SV, structural validity; IC, internal consistency; Rel, reliability; ME, measurement error; CCV, cross-cultural validity; CrV, criterion validity; HT, hypothesis testing; Resp, responsiveness; App, application; Acc, acceptability.

Appendix B: Inclusion and exclusion criteria

Criterion	Requirement	Rationale
1. Construct alignment	Instruments must directly measure a construct defined within the theoretical taxonomy; tools assessing only biological targets were excluded	<i>Ensures conceptual consistency and prevents inclusion of measures unrelated to contextual meaning</i>
2. Psychometric validation	Only instruments with peer-reviewed validation evidence (e.g., factor analysis, internal consistency) were included; unvalidated checklists were excluded	<i>Maintains psychometric rigor and supports reliability and validity of included measures</i>
3. Originality	Original, validated instruments were prioritized; modified, translated, or abbreviated versions were excluded unless the adapted version is the accepted clinical standard	<i>Reduces variance introduced by derivative tools and preserves measurement fidelity</i>
4. Focus on contextual throughputs	Measures of stable patient traits or phenotypes unrelated to the clinical encounter (e.g., Locus of Control, Patient Activation Measure, Functional Independence Measure, Jefferson Empathy patient inventories) were excluded. Baseline psychological traits known to moderate placebo/nocebo mechanisms per Bishop's taxonomy were retained	<i>Retains focus on contextual factors that directly influence or frame immediate healthcare delivery rather than isolated patient predispositions</i>

Acknowledgements This work was funded by the National Center for Complimentary and Integrative Health (NCCIH) and the National Institute of Neurological Disorders and Stroke (NINDS), U24AT01196 to WRR and CC. The funder had no role in the design and the conduct of the study, collection, management, analysis and interpretation of the data, preparation, review or approval of the manuscript, or decision to submit the manuscript.

Author contributions CC and GR conceived and designed the study. CC, JB and GR and provided ongoing coordination of the expert panel. All authors contributed to the expert identification, appraisal, and selection of instruments. JB, GR, and CC performed the data synthesis, gap analysis, and instrument mapping. JB created the images and final tables for the manuscript. All authors participated in drafting the manuscript, provided critical intellectual content through revisions, and approved the final version of the manuscript for submission.

Funding This work was supported by NIH U24 Force-Based Manipulation Network (1U24AT011969) from the National Center for Complementary and Integrative Health & National Institute of Neurological Disorders and Stroke (NINDS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability The deidentified quantitative psychometric scores, standardized evaluation matrices, and the full multi-criteria decision analysis (MCDA) dataset generated during this study, as well as the primary evaluation framework and domain taxonomies, are fully disclosed within this manuscript and its supplementary files. To ensure full transparency and immediate clinical utility, the registry of all 57 psychometrically robust instruments evaluated by the consensus panel is securely archived and publicly accessible via the Open Science Framework (OSF) at: <https://osf.io/jf2sg/overview>

Code availability No custom computer code or mathematical algorithms were used to generate the results in this study.

Competing interests JB, CC, VN, WRR have a portion of their salary funded by ForceNET U24AT01196. JB reported holding an adjunct faculty appointment at Medical College of South Carolina, CC reported serving as the Director for the Center of Excellence in Manual and Manipulative Therapy at Duke University with a portion of his supported by that role; receiving royalties for a textbook on OMT and a course on manual therapy with AGENCE EBP; and serving as a consultant

with the Hawkins Foundation, Zimmer Biomedical and Akron Children's. VN reported serving as a paid consultant for Cala Health. JE reported serving as Associate Editor for the International Journal of Osteopathic Medicine. GR reported leading educational and continuing professional development courses (CPD) on placebo, nocebo effects and CFs in healthcare. No other disclosures were reported by the other authors.

Ethics approval and consent to participate This study was reviewed by the Duke University Health System Institutional Review Board and was determined to be exempt human subjects research (Protocol Pro00117707). The need for informed consent was waived by the Duke University Health System IRB under this protocol. All methods were carried out in accordance with relevant institutional policies, federal regulations, and applicable guidelines for research involving human participants.

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/>.

References

1. Newell, D. et al. Prediction, pain, and the power of context: Toward a contemporary identity for the chiropractic profession. *Chiropr. Man. Ther.* **34**, 35 (2026).
2. Cook, C., Young, C., Harvey, D. & Rossettini, G. Beyond treatment: How care process factors shape physical therapist management of low back pain. *Physiother. Theory Pract.* <https://doi.org/10.1080/09593985.2026.2716786> (2026).
3. Ezzatvar, Y., Dueñas, L., Balasch-Bernat, M., Lluch-Girbés, E. & Rossettini, G. Which portion of physiotherapy treatments' effect is not attributable to the specific effects in people with musculoskeletal pain? A meta-analysis of randomized placebo-controlled trials. *J. Orthop. Sports Phys. Ther.* **54**, 391–399 (2024).
4. Saueressig, T., Pedder, H., Owen, P. J. & Belavy, D. L. Contextual effects: How to, and how not to, quantify them. *BMC Med. Res. Methodol.* **24**, 35 (2024).
5. Hohenschurz-Schmidt, D. et al. Placebo analgesia in physical and psychological interventions: Systematic review and meta-analysis of three-armed trials. *Eur. J. Pain.* **28**, 513–531 (2024).
6. Mokkink, L. B. et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med. Res. Methodol.* **10**, 22 (2010).
7. Prinsen, C. A. C. et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res.* **27**, 1147–1157 (2018).
8. Fazey, I. & Fazey, J. A. Consensus development panel. In *Strengthening Transdisciplinary Research: Integration and Implementation Sciences for Tackling Complex Problems* (ed. Bammer, G.) (ANU Press, 2013).
9. Cook, C. E. et al. An international consensus definition for contextual factors: Findings from a nominal group technique. *Front. Psychol.* **14**, 1178560 (2023).
10. Bishop, F. L. et al. What techniques might be used to harness placebo effects in non-malignant pain? A literature review and survey to develop a taxonomy. *BMJ Open* **7**, e015516 (2017).
11. Arksey, H. & O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **8**, 19–32 (2005).
12. Williamson, P. R. et al. The COMET handbook: Version 1.0. *Trials* **18**(Suppl. 3), 280 (2017).
13. Prinsen, C. A. et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" a practical guideline. *Trials* **17**, 449 (2016).
14. Thokala, P. et al. Multiple criteria decision analysis for health care decision making an introduction. *Value Health.* **19**, 1–13 (2016).
15. Marsh, K. et al. Multiple criteria decision analysis for health care decision making emerging good practices. *Value Health* **19**, 125–137 (2016).
16. University of Bath Research Portal. Lexicographic decision rule. *University of Bath Research Portal* (University of Bath, n.d.).
17. Scheier, M. F., Carver, C. S. & Bridges, M. W. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* **67**, 1063–1078 (1994).

18. Schwarzer, R. & Jerusalem, M. Generalized self-efficacy scale. In *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs* (eds Johnston, M. & Wright, S.) 35–37 (NFER-NELSON, 1995).
19. Kalauokalani, D. et al. Lessons from a trial of acupuncture and massage for low back pain: Patient expectations and treatment outcomes. *Spine* **26**, 1418–1424 (2001).
20. Shedden-Mora, M. et al. The Treatment Expectation Questionnaire (TEX-Q): Validation of a generic multidimensional scale measuring patients' treatment expectations. *BMJ Open* **13**, e067602 (2023).
21. Mayer, T. G. et al. The development and psychometric validation of the central sensitization inventory (CSI). *Pain Pract.* **12**, 276–285 (2012).
22. Giordano, L. A. et al. Development, implementation, and public reporting of the HCAHPS survey. *Med. Care Res. Rev.* **67**, 27–37 (2010).
23. Williams, D. R. et al. Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *J. Health Psychol.* **2**, 335–351 (1997).
24. Mehling, W. E. et al. The multidimensional assessment of interoceptive awareness (MAIA). *PLoS ONE* **13**, e0208034 (2018).
25. Silva, A. C. et al. Psychometric properties of instruments to evaluate evidence-based practice: a systematic review. *Health Educ. Behav.* (2024).
26. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).
27. Hall, M. A. et al. Measuring patients' trust in their primary care providers. *Med. Care Res. Rev.* **59**, 293–318 (2002).
28. Carron, A. V., Widmeyer, W. N. & Brawley, L. R. The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. *J. Sport Psychol.* **7**, 244–266 (1985).
29. Richmond, V. P., McCroskey, J. C. & Johnson, A. D. Development of the Nonverbal Immediacy Scale (NIS): Measures of self- and other-perceived nonverbal immediacy. *Commun. Q.* **51**, 504–517 (2003).
30. Burnell, L. & Agan, D. L. Compassionate care: Can it be defined and measured? The development of the Compassionate Care Assessment Tool (CCAT). *Int. J. Caring Sci.* **6**, 180–187 (2013).
31. Mercer, S. W., Maxwell, M., Heaney, D. & Watt, G. C. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam. Pract.* **21**, 699–705 (2004).
32. Garvelink, M. M. et al. Psychometric properties of the decisional conflict scale: A systematic review. *Med. Decis. Making* **39**, 544–559 (2019).
33. Gremigni, P. et al. The health care communication questionnaire (HCCQ-Comm). *Health Commun.* (2016).
34. Eisenberger, R. et al. Perceived organizational support. *J. Appl. Psychol.* **71**, 500–507 (1986).
35. Zagreus, L. et al. Listening to the occupants: A web-based indoor environmental quality survey. *Indoor Air* **14**(Suppl. 8), 65–74 (2004).
36. Borkovec, T. D. & Nau, S. D. Credibility of analogue therapy rationales. *J. Behav. Ther. Exp. Psychiatry* **3**, 257–260 (1972).
37. Moss-Morris, R. et al. The revised illness perception questionnaire (IPQ-R). *Psychol. Health* **17**, 1–16 (2002).
38. Zigmond, A. S. & Snaith, R. P. The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* **67**, 361–370 (1983).
39. Houben, R. M. et al. Health care providers' orientations towards common low back pain predictions for back pain related disability. *Pain* **114**, 491–498 (2005).
40. Axelsson, Ö., Nilsson, M. E. & Berglund, B. A principal components model of soundscape perception. *J. Acoust. Soc. Am.* **128**, 2836–2846 (2010).
41. Doty, R. L., Shaman, P. & Dann, M. Development of the University of Pennsylvania smell identification test: A standardized microencapsulated test of olfactory function. *Physiol. Behav.* **32**, 489–502 (1984).
42. Hummel, T., Sekinger, B., Wolf, S. R., Pauli, E. & Kobal, G. "Sniffin" sticks': Olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chem. Senses* **22**, 39–52 (1997).
43. Lee, J. J. et al. Development and psychometric validation of the olfactory dysfunction outcomes rating. *JAMA Otolaryngol. Head Neck Surg.* **148**, 1132–1139 (2022).
44. Makoul, G., Krupat, E. & Chang, C. H. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool (CAT). *Patient Educ. Couns.* **67**, 333–342 (2007).
45. Roter, D. & Larson, S. The Roter interaction analysis system (RIAS): Utility and flexibility for analysis of medical interactions. *Patient Educ. Couns.* **46**, 243–251 (2002).
46. Wilhelm, F. H., Kochar, A. S., Roth, W. T. & Gross, J. J. Social anxiety and response to touch: Consonant and dissonant findings. *J. Anxiety Disord.* **15**, 279–294 (2001).
47. Department of Health Estates and Facilities Division. *ASPECT: A Staff and Patient Environment Calibration Toolkit* (Department of Health, 2008).
48. Vafeiadou, A., Bowling, N. C., Hammond, C. & Banissy, M. D. Assessing individual differences in attitudes towards touch in treatment settings: introducing the Touch & Health Scale. *Health Psychol. Open* **9**, 20551029221137056 (2022).

49. Rossettini, G., Palese, A. & Cook, C. Trying to explain the unexplainable": Why research on contextual factors in musculoskeletal pain is needed. *Pain Manag.* **14**, 465–468 (2024).
50. Poulter, D. et al. Contextual effects in musculoskeletal pain: Are we overlooking essential factors?. *Front. Psychol.* **16**, 1537242 (2025).
51. Hojat, M. et al. Physician empathy: Definition, components, measurement, and relationship to gender and specialty. *Am. J. Psychiatry* **159**, 1563–1569 (2002).
52. Langevin, H. M. Moving the complementary and integrative health research field toward whole person health. *J. Altern. Complement. Med.* **27**, 623–626 (2021).
53. Hutting, N. Self-management support is not optional in musculoskeletal care. *Musculoskelet. Sci. Pract.* **84**, 103597 (2026).
54. Laursen, J., Danielsen, A. & Rosenberg, J. Effects of environmental design on patient outcome: A systematic review. *HERD* **7**, 108–119 (2014).
55. Rowe, A. & Knox, M. The impact of the healthcare environment on patient experience in the emergency department: A systematic review to understand the implications for patient-centered design. *HERD* **16**, 310–329 (2023).
56. Kaptchuk, T. J. Placebo studies and ritual theory: A comparative analysis of Navajo, acupuncture and biomedical healing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 1849–1858 (2011).

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

Authors and Affiliations

Jennifer A. Bent^{1,2}  · Chad E. Cook^{3,4,5} · Antoine Bailliard⁶ · Joel E. Bialosky^{7,8} · Felicity L. Bishop⁹ · Elisa Carlino^{10,11} · Luana Colloca^{12,13,14} · Jorge E. Esteves¹⁵ · Alessandra Garcia¹⁶ · Amy W. McDevitt¹⁷ · Vitaly Napadow¹⁸ · David Newell¹⁹ · Alvisa Palese²⁰ · William R. Reed²¹ · Giacomo Rossettini^{19,22,23}

✉ Jennifer A. Bent
Jennifer.bent@duke.edu

- ¹ Department of Rehabilitation Services, Duke University Health Care System, Durham, 5601 Arrington Park Drive, Suite 203, Morrisville, NC 27513, USA
- ² Department of Rehabilitation Sciences, Medical University of South Carolina, Charleston, SC, USA
- ³ Department of Orthopaedics, Duke University, Durham, NC, USA
- ⁴ Duke Clinical Research Institute, Duke University, Durham, NC, USA
- ⁵ Department of Population Health Sciences, Duke University, Durham, NC, USA
- ⁶ Occupational Therapy Doctorate Division, Department of Orthopaedic Surgery, Duke University, Durham, NC, USA
- ⁷ Department of Physical Therapy, University of Florida, Gainesville, FL, USA
- ⁸ Clinical Research Center, Brooks Rehabilitation, Jacksonville, FL, USA
- ⁹ School of Psychology, University of Southampton, Southampton, Hampshire, UK
- ¹⁰ “Rita Levi Montalcini” Department of Neurosciences, University of Turin, Torino, Italy
- ¹¹ U.O. Di Neurologia E Neuroriabilitazione, I.R.C.C.S. Istituto Auxologico Italiano, Ospedale San Giuseppe, Piancavallo (VCO), Italy
- ¹² Department of Pain and Translational Symptom Science, School of Nursing and Department of Anesthesiology School of Medicine, University of Maryland, Baltimore, MD, USA
- ¹³ Placebo Beyond Opinions Center, University of Maryland School of Nursing, Baltimore, MD, USA

- ¹⁴ College of Computer, Mathematical and Natural Sciences, University of Maryland College Park, College Park, MD, USA
- ¹⁵ Escola Superior de Saúde Atlântica, Barcarena, Portugal
- ¹⁶ College of Pharmacy & Health Sciences, Campbell University, Lillington, NC, USA
- ¹⁷ Department of Physical Medicine and Rehabilitation, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA
- ¹⁸ Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation, Harvard Medical School, Boston, MA, USA
- ¹⁹ Centre for Pain and Active Inference Research, Health Sciences University, Bournemouth, UK
- ²⁰ Department of Medicine, University of Udine, Udine, Italy
- ²¹ Department of Physical Therapy, University of Alabama at Birmingham, Birmingham, AL, USA
- ²² School of Physiotherapy, University of Verona, Verona, Italy
- ²³ Department of Physiotherapy, Faculty of Sport Sciences, Universidad Europea de Madrid, Villaviciosa de Odon, Spain