Development of a Brief Multidimensional Index of a Rehabilitation Clinician's Propensity to Integrate Research Evidence into Clinical Decision-Making

Jacqueline Roberge-Dao, MSc. A. (Occupational Therapy)
School of Physical and Occupational Therapy
Faculty of Medicine
McGill University
Montreal, Canada

August 2022

A thesis submitted to the Faculty of Graduate Studies and Research, McGill University in partial fulfillment for the requirements of the degree of

DOCTOR OF PHILOSOPHY

(Rehabilitation Sciences)

McGill

Table of Contents

List of tables

List of figures

List of appendices

Abbreviations

ABBREVIATION	MEANING
BWS	Best-worst scaling
CDM	Clinical decision-making
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
CREATE	Classification Rubric for EBP Assessment Tools in Education
DF	Degrees of freedom
DIF	Differential Item Functioning
EBM	Evidence-based medicine
EBP	Evidence-based practice
ERIC	Education Resources Information Center
GEE	Generalized estimating equation
I-TIDP-PDC	*Indice de la tendance à intégrer les données probantes dans la prise de décisions cliniques*
ICC	Item characteristic curve
ICF	International Classification of Functioning, Disability and Health
JBI	Joanna Briggs Institute
K	Kappa
KT	Knowledge translation
MEDLINE	Medical Literature Analysis and Retrieval System Online
N	Number
N/A	Not applicable
NB	*Nota bene*
OT	Occupational therapist
P-PIRE-CDMI	Prototype of the Propensity to Integrate Research Evidence into Clinical Decision-Making Index
PBM	Preference-based measure
PFDP	*Pratique fondée sur les données probantes*
PIRE-CDMI	Propensity to Integrate Research Evidence into Clinical Decision-Making Index
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient-reported out- come measure
PROSPERO	The International Prospective Register of Systematic Reviews
PSI	Person Separation Index
PT	Physical therapist
R	Correlation coefficient
RCT	Randomized controlled trial

RLH	Root-likelihood
RMT	Rasch Measurement Theory
SD	Standard Deviation
SE	Standard Error
SLR	Simple linear regression
SR	Systematic review
SRM	Standardized response mean
THE STANDARDS	The Standards of Psychological and Educational Testing
VAS	Visual analogue scale
ß	Beta

Abstract

Evidence-based practice (EBP) is considered a priority in health care and is now a professional requirement for Canadian occupational (OT) and physical therapists (PT). EBP is defined as a complex and multidimensional approach to clinical decision-making (CDM) that combines the best available research evidence, clinical expertise, and patients' preferences, within the context of available resources. Despite the anticipated benefits of EBP and the growing emphasis on this approach to CDM, engagement of OTs and PTs with the EBP process remains a challenge.

Robust measurement practices are essential to determine if clinicians are engaging in the EBP process and to identify the factors related to EBP that should be improved or maintained. The measurement of EBP can serve to identify research and practice needs, better support clinicians and ultimately, improve quality of care. One approach to measuring EBP is through self-report questionnaires composed of one or multiple EBP domains known to influence the enactment of EBP. Six core domains have been previously identified and represent individual and organizational constructs that influence EBP such as attitudes towards evidence, self-efficacy in applying evidence to practice or participation in activities related to EBP. Despite important developments in the measurement of EBP, there are still challenges including (1) the shortage of comprehensive questionnaires representing the most salient individual and organizational domains; (2) the inappropriate treatment of ordinal data as interval-level data; (3) the fragmented interpretation of EBP due to "profile-type" measures which summarize results of domains independently; and (4) the unknown relative importance of the domains that influence EBP and how these can be quantitatively captured in a measure.

The overall objective of this doctoral thesis was to take the necessary steps towards developing a brief, multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into CDM to use as a global outcome in clinical research.

To operationalize the overarching aim of my doctoral work, I conducted four research projects (described in five manuscripts that compose this thesis).

In the first study, I carried out an umbrella review which aimed to describe the methods, results, recommendations, reported challenges and areas for future practice across systematic reviews on EBP measures in healthcare. Findings highlighted the scarcity of comprehensive measures covering multiple domains and the lack of rigor in the development and testing of

measures (Manuscripts 1 and 2). Findings from this review reinforced the need for a robust and comprehensive measure of EBP and informed the process for the next phase.

In the second study, I carried out the initial steps required to develop the Propensity to Integrate Research Evidence into Clinical Decision-Making Index (PIRE-CDMI). Expanding on earlier research, I reduced an initial pool of 70 items covering six EBP domains to five items using previously conducted Rasch analysis and expert feedback (Manuscript 3). I developed a preliminary scoring algorithm which generated a single score for testing purposes and examined the interpretability of the prototype index scores across characteristics of the sample of OTs and PTs and compared to other EBP measures (from a previous dataset). Findings supported the continuation of the PIRE-CDMI development process.

The five items included in the prototype index were derived from various questionnaires resulting in differences in terminology and formulation of items and response levels which can increase respondent burden and measurement bias. For the third study, I conducted a focus group and cognitive interviews with target users of the measure (OTs and PTs) to contribute evidence for the clarity and interpretability of items and response levels in English and French. Important modifications were made to the items based on participant feedback. Additionally, I carried out a cross-sectional online survey to demonstrate the equivalency of response labels in both languages (Manuscript 4).

The final study in this dissertation describes the procedures required to develop a weighted scoring algorithm for the index. I used specialized software and launched an online survey to elicit part-worth utilities for all 15 dimension-levels (five items, three response levels per item) from the perspective of OTs, PTs, clinical managers, and experts in EBP (Manuscript 5). A choice exercise, called best-worst scaling, was used to estimate the part-worth utilities which were integrated into an algorithm allowing for an overall score to be representative of the gains in one domain against the losses in another. This study also provides strong evidence regarding the interpretability of index scores with OTs and PTs.

The original contribution of this doctoral research is the development of the first weighted multidimensional index in rehabilitation which generates a mathematically and theoretically valid total score of a rehabilitation clinician’s propensity to integrate research evidence into clinical decision-making.

Abrégé

La pratique fondée sur les données probantes (PFDP) est une priorité dans le secteur de la santé et constitue désormais une obligation professionnelle pour les ergothérapeutes et physiothérapeutes. La PFDP est définie comme une approche complexe et multidimensionnelle à la prise de décisions cliniques qui combine les données probantes, l'expertise clinique et les préférences des patients, dans le cadre des ressources disponibles. Malgré les avantages anticipés et l'emphase mise sur la PFDP, les ergothérapeutes et physiothérapeutes continuent à avoir des difficultés à mettre en œuvre la PFDP.

Des outils de mesure sont nécessaires pour déterminer si les cliniciens participent à la PFDP et pour identifier les éléments liés à la PFDP qui devraient être améliorés ou maintenus. Mesurer la PFDP permet d'identifier les besoins en matière de recherche et de pratique, de mieux soutenir les cliniciens et, d'améliorer la qualité des services. Une façon de mesurer la PFDP consiste à utiliser des questionnaires d'auto-évaluation, qui doivent être complétés par les cliniciens et qui sont composés d'un ou de plusieurs domaines influençant la PFDP. Six domaines ont été identifiés comme étant des facteurs clés contribuant à la PFDP et représentent des construits individuels et organisationnels; il y a par exemple, les attitudes envers les données probantes, l'auto-efficacité dans l'application des données probantes en pratique, ou encore, la participation à des activités liées à la PFDP. Malgré d'importants développements en ce qui trait aux outils de mesure de la PFDP, certains défis demeurent, notamment : (1) le manque de questionnaires représentant de manière compréhensive les facteurs clés influençant la PFDP; (2) le fait que les données ordinales soient traitées comme des données d'intervalles ; (3) la limitation des questionnaires existants de par leur nature de type « profil » résument les résultats de sous-échelles (ou domaines) de manière isolée; et (4) la méconnaissance de l'importance relative des domaines qui influencent la PFDP et de comment ceux-ci peuvent être capturés de manière quantitative dans une mesure.

L'objectif global de cette thèse doctorale est d'entreprendre les étapes nécessaires pour le développement d'un bref indice multidimensionnel qui permettra de mesurer la tendance d'un clinicien en réadaptation à intégrer les données probantes dans la prise de décisions cliniques pour être utilisé comme un résultat global dans la recherche clinique.

Pour mettre en œuvre l'objectif global de mon doctorat, j'ai mené quatre projets de recherche (décrits dans cinq manuscrits qui composent cette thèse).

Tout d'abord, j'ai effectué une revue des revues systématiques (appelé «umbrella review» en anglais) qui avait pour but de décrire les méthodes, les résultats, les recommandations, les défis rapportés et les futurs champs de pratique parmi les revues systématiques sur les outils de mesure de la PFDP. Les résultats ont mis en évidence la rareté des outils de mesure couvrant de manière compréhensive les domaines clés influençant la PFDP et le manque de rigueur dans le développement des outils de mesure (*Manuscrits 1 et 2).* Les résultats de cette revue ont renforcé le besoin de développer un outil de mesure robuste et multidimensionnelle de la PFDP et ont informé la prochaine phase de mon projet.

Dans la deuxième étude, j'ai entrepris les démarches préalables au développement de *l'indice de la tendance à intégrer les données probantes dans la prise de décisions cliniques* (I-TIDP-PDC). En me fondant sur des recherches antérieures, j'ai réduit un groupe de 70 items couvrant six domaines liés à la PFDP à cinq items en utilisant l'analyse Rasch (précédemment réalisée par d'autres chercheurs) et les avis d'experts (*Manuscrit 3*). Ensuite, j'ai développé un système de pointage préliminaire qui a permis de générer un score afin de tester les résultats de l'indice prototype avec l'échantillon d'ergothérapeutes et de physiothérapeutes (à partir d'un ensemble de données existant). Des preuves de validité pour sa structure interne et ses relations avec d'autres variables ont appuyé la poursuite du processus de développement de l'indice.

Les cinq items inclus dans l'indice prototype ont été dérivés de divers questionnaires, ce qui entraîne des différences dans la terminologie et formulation des items et des choix de réponse, pouvant augmenter le travail des répondants et le biais de mesure. Pour la troisième étude, j'ai donc mené un groupe de discussion et des entretiens cognitifs avec des utilisateurs cibles de l'outil afin de vérifier la clarté et l'interprétabilité des items et des choix de réponse en anglais et en français. Des modifications importantes ont été apportées aux items en fonction des commentaires reçus des participants. De plus, j'ai réalisé un sondage en ligne pour démontrer l'équivalence des choix de réponses dans les deux langues (*Manuscrit 4*).

La dernière étude de cette thèse décrit l'élaboration d'un système de pointage pour l'indice. J'ai employé un logiciel spécialisé et lancé un sondage en ligne pour obtenir les poids pour les 15 niveaux de réponse (cinq items, trois choix de réponse par item) du point de vue des ergothérapeutes, des physiothérapeutes, des gestionnaires cliniques et des experts (*Manuscrit 5*). Un exercice de choix, appelé « le meilleur et le pire choix », a été utilisé pour estimer les poids qui ont ensuite été intégrés dans un algorithme permettant de produire un score global représentatif des acquis dans un domaine par rapport aux déficits dans un autre. Cette étude fournit également des preuves de validité concernant l'interprétabilité des scores de l'indice auprès des ergothérapeutes et des physiothérapeutes.

La contribution originale de cette recherche doctorale est l’élaboration du premier indice multidimensionnel en réadaptation qui produit un score global mathématiquement et théoriquement valide de la tendance d'un clinicien en réadaptation à intégrer les données probantes dans la prise de décisions cliniques.

Acknowledgements

This dissertation would not have been possible without the support of several people. To my supervisors, committee members, professors, colleagues, friends, family, and to everyone with whom I have collaborated along this PhD journey, I am eternally grateful for your teachings and support over the past few years. This PhD is a cumulation of learnings from each of you- thank you!

Firstly, I would like to express my deepest gratitude to Dr. Aliki Thomas for her unwavering support and expert guidance throughout the entire PhD journey. Thank you for introducing me to the world of academia. It has been an absolute privilege to learn and discuss innovative ideas with such a distinguished scholar. You have modeled what it is to be a brilliant, conscientious, and graceful researcher. You always go the extra mile for your students, and I will forever be grateful for your mentorship.

I would also like to extend my sincere gratitude to Dr. Nancy Mayo. It has been an incredible honor and gift to learn from you. You are the epitome of a great educator - your enthusiasm for “asking the right question” and “getting closest to the right answer” is contagious. I will always remember you saying, “data is best served raw, like veggies”. You have helped me in building a comprehensive methodological toolbox that I will carry with me forever.

I wish to also thank my committee members for their incredible guidance and generous feedback in preparing this thesis. Dr. Annie Rochette, you have been an important pillar throughout my PhD journey. Thank you for your thought-provoking questions, attention to detail, and for being a source of zenitude and conviviality. Dr. Keiko Shikako, thank you for your consistent enthusiasm and sound advice on academic writing and on how to deeply consider the implications of the research that we do. It has been an honor to learn alongside you both and I am fortunate that our paths have crossed.

Dear members of the *KEEP* and *OutcomesRUs* labs, I am thankful to every one of you who has helped with my research and who has lent attentive ears. I send a special thanks to Marco Zaccagnini and Rebecca Ataman for your support and for the stimulating conversations about life inside and outside the lab. Marie-Christine Hallé, Susanne Mak, Lisa Arcobelli, Cathy George, Catherine Giroux, Conchita Saldanha, Sungha Kim, Swati Das, Jennifer Fitzpatrick and Zafar Iqbal- I hold very fond memories of working alongside you in the lab. Thank you Fadi Alzoubi for explaining to me your important Rasch analysis work at the start of this research project. Thanks to Ana Maria Moga and Kedar Mate with whom I have shared great discussions. Fatima, Rima, Sarah, and Carly- thank you for always being sources of kindness and support. Thank you, Lauren Maggio and Jill Boruff, for your expertise and support.

I would also like to express my immense gratitude to all the participants who took part in this project and made this doctoral research possible. This thesis would not have been possible without the financial support of the Richard and Edith Strauss Foundation, the School of Physical and Occupational Therapy at McGill University, the Faculty of Medicine and Health Sciences at McGill University, the Centre de Recherche Interdisciplinaire en Réadaptation du Montréal Métropolitain and Société Inclusive.

To my family and friends, I would like to share this achievement with you. Your continued love, support and patience have kept me going in the most difficult times. To my sister, Julia, you are simply the best; your surprise notes scattered in my office have given me strength. To my parents - the strong principles that you have instilled in me are present in everything that I do. I am eternally thankful for your unconditional love and support. I also wish to recognize those who have paved the way for me, often with lots of sacrifice.

Last, but not least- to Mathieu. I am forever indebted to you for your inexhaustible moral support and love. Thank you for standing by me through all the highs and lows of this PhD journey. Thank you for always believing in me, even when at times I didn't believe in myself.

Preface

Statement of originality

This thesis describes the steps taken towards developing a brief multidimensional index of a clinician's propensity to integrate research evidence into clinical decision-making. I first became interested in the measurement of evidence-based practice (EBP) and knowledge translation (KT) during my master's professional project which focused on the evaluation of a KT funding initiative (53 research projects) at the School of Physical and Occupational Therapy at McGill University. Participating in that research project provided me with a broad understanding of existing KT research and shed light on the strengths and shortcomings of implementation efforts. Then, during my first year as a practicing occupational therapist, I became increasingly aware of the disconnect between what clinicians do and the best available evidence. Like many colleagues, I experienced barriers in my own practice to using scientific evidence to inform my decisions about the care of patients. The lack of organizational support and regular monitoring regarding the integration of research evidence into practice made being an evidence-based clinician challenging for me. I felt like EBP was this abstract ideal and although I was doing my best to use research evidence, I never knew how I could concretely improve my ability to do so. This frustration and curiosity fueled my desire to pursue doctoral studies.

I started off the PhD journey with a deep interest in the measurement of EBP stemming from the above-mentioned challenge and my interest in measurement and evaluation. The idea of developing a brief index related to EBP arose from a group discussion during my first committee meeting. My supervisor (Aliki Thomas) and co-supervisor (Nancy Mayo) were jointly involved in a Canadian Institutes of Health Research funded longitudinal project examining the evolution in EBP among newly graduated occupational and physical therapists across Canada. Coincidentally, they had just completed the development of a measure using Rasch analysis which was led by a post-doctoral fellow (Fadi Alzoubi). In line with my interests, my co-supervisor highlighted the possibility of expanding on the work that had just been conducted (Rasch analysis) by working towards the development of a brief index using the general steps used to develop preference-based measures.

The original contribution of this thesis lies in the development of the first weighted multidimensional index of propensity to integrate research evidence into decision-making. In doing so, this thesis makes important theoretical, methodological, and practical contributions to the field of EBP, specifically in rehabilitation.

Although this research could not have been conducted without the guidance from my supervisor Aliki Thomas, my co-supervisor Nancy Mayo, and my supervisory committee Annie Rochette and Keiko Shikako, this statement attests that the writing and manuscripts presented in this doctoral dissertation are my own original work. I confirm that any assistance received in carrying out this research has been acknowledged. This dissertation has not been submitted elsewhere, apart from *Manuscripts 1 and 2* which are published in scientific journals. Please note that an external editor has not been used in the construction of this thesis.

Contribution of authors

The manuscripts in this dissertation are the work of doctoral candidate Jacqueline Roberge-Dao with guidance from supervisor Aliki Thomas (PhD) and co-supervisor Nancy Mayo (PhD). For all five manuscripts, the study design, ethics application (when relevant), data collection (apart from *Manuscript 3* which used an existing dataset), statistical analysis, interpretation of results and manuscript writing were conducted by the doctoral candidate with guidance from Aliki Thomas and Nancy Mayo. For *Manuscripts 4 and 5*, the doctoral candidate was also responsible for recruiting participants for a focus group, cognitive interviews, and two cross-sectional surveys. The doctoral candidate was responsible for the originality of the ideas, under the advisement of Aliki Thomas and Nancy Mayo, the scientific quality of the research and for the quality of the reporting across all manuscripts in this dissertation.

As primary supervisor, Aliki Thomas oversaw all aspects of the thesis, contributed her expertise regarding the theory and methods related to measuring EBP and provided extensive editing and feedback on all chapters of this thesis. She is also co-author on all five manuscripts.

As co-supervisor, Nancy Mayo oversaw all aspects of the thesis, provided expertise regarding research methodology and statistical analysis, and provided feedback on all chapters of this thesis. She is also co-author on *Manuscripts 3, 4, and 5*.

As committee members, Annie Rochette (PhD) and Keiko Shikako (PhD) provided content and editorial feedback on all chapters of this thesis and are co-authors for all five manuscripts.

Lauren Maggio (PhD) was a co-author in *Manuscripts 1 and 2* for her feedback as to the methodological implications of conducting an umbrella review, specifically the pitfalls to avoid given her experience with a previous umbrella review. She also provided guidance with data analysis given her expertise in EBP and conducting syntheses. She also provided editorial feedback on these manuscripts.

Marco Zaccagnini (PhD candidate) was a co-author in *Manuscripts 1 and 2* as he was the second reviewer for the umbrella review search, study selection, data extraction and critical appraisal process. He also provided editorial feedback on these manuscripts.

Jill Boruff was a co-author in *Manuscripts 1 and 2* as the health sciences librarian who assisted with the search strategy. She also provided editorial feedback on these manuscripts.

Thesis organization and overview

This dissertation consists of five manuscripts, two of which have already been published in peer-reviewed scientific journals. In alignment with the *Graduate and Postdoctoral Studies* regulations at McGill University, an introduction chapter, bridging chapters between the five manuscripts and a discussion chapter have been incorporated in this dissertation. For this reason, some information may be repeated throughout this thesis. The following consists of a brief outline of this thesis:

Chapter 1 consists of the introduction to the doctoral work; it contains the results of a literature review that was conducted to present and describe the main topics comprised in this thesis including a historical overview of the EBP movement; the state of EBP in rehabilitation; and the measurement of EBP in rehabilitation.

Chapter 2 presents the rationale for this research and the objectives of each Manuscript.

Chapter 3.1 consists of *Manuscript 1*, entitled *Quality, Methods, and Recommendations of Systematic Reviews on Measures of Evidence-Based Practice: An Umbrella Review*. This is the first part of a comprehensive umbrella review on systematic reviews of EBP measures.

Chapter 3.2 consists of *Manuscript 2*, entitled *Challenges and Future Directions in the Measurement of Evidence-based Practice: Qualitative Analysis of Umbrella Review Findings*. This is the second part of the previously described umbrella review.

Chapter 4 links the first and second manuscripts to the third manuscript.

Chapter 5 consists of *Manuscript 3*, entitled *Identifying Candidate Items for a Prototype Index on Propensity to Integrate Research Evidence into Clinical Decision-making in Rehabilitation*, describes the prototype development of the index and provides evidence to support the continued development of this index.

Chapter 6 links the third manuscript to the fourth manuscript.

Chapter 7 consists of *Manuscript 4*, entitled *Improving the Clarity and Interpretability of Items in a Bilingual Index of Propensity to Integrate Research Evidence into Clinical Decision-making in Rehabilitation*, and describes the revision and rewriting of items included in the prototype index for clarity and interpretability of items in English and French.

Chapter 8 links the fourth manuscript to the fifth manuscript.

Chapter 9 consists of *Manuscript 5*, entitled *Propensity to Integrate Research Evidence into the Decision-making Process in Rehabilitation: Development of a Weighted Algorithm Using a Best-Worst Scaling Choice Exercise*, and describes how part-worth utilities were elicited for each dimension and integrated into a scoring algorithm.

Chapter 10 presents the discussion chapter which includes a summary of the main findings of this doctoral research; the theoretical, methodological, and practical contributions of this research; strengths and limitations; areas of future research; and a concluding statement.

Tables and figures are presented within each manuscript. References and appendices for all chapters can be found after each chapter. Reference styles were based the *American Medical Association (AMA) Manual of Style*, except for Chapter 3.1 which followed the journal's requirements. All projects requiring ethics approval were approved by *The Faculty of Medicine and Health Sciences Institutional Review Board* at McGill University and all participants provided informed consent.

Chapter 1: Literature review

Healthcare clinicians including physicians, nurses, occupational therapists (OTs) and physical therapists (PTs) make several decisions daily related to the direct and indirect care of patients, logistics (e.g., scheduling) and resource management. Clinical decision-making (CDM) is the "ability to sift and synthesize information, make decisions and appropriately implement these decisions in the clinical environment"[1]. With the mission of optimizing health, well-being and quality of life, clinicians must draw on different sources of knowledge including research evidence, professional experience, physiological or basic scientific knowledge, patient goals and values, and system features[2] to make important decisions about diagnosis, prediction, classification, and treatment[3]. One approach that has been highly advocated in the last three decades for informing CDM and improving the quality of clinical practice is evidence-based practice[4].

Historical overview of the evidence-based practice movement

EBM models

In a 1991 editorial, Gordon Guyatt introduced evidence-based medicine (EBM) to the medical community as a new approach to medical practice which "requires skills of literature retrieval, critical appraisal, and information synthesis" [5(pA16)]. A year later, a highly cited *JAMA* paper was published in which scholars from McMaster University including David Sackett and Gordon Guyatt declared that "a new paradigm for medical practice is emerging" and stressed the development of new skills for the "examination of evidence from clinical research"[6(p2420)] . EBM was defined as "the process of systematically finding, appraising, and using contemporaneous research findings as the basis for clinical decisions" [6(p2420)]. EBM emphasized the primacy of systematic and reproducible observations derived from high-quality trials over expert opinion, intuition, and physiological reasoning[6].

Four years later in 1996, the pioneers of the EBM movement updated the definition to: "EBM is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research."[7(p71)]. This update explicitly defined EBM as a model for CDM with three separate yet interconnected components: patient preferences, clinical expertise, and research evidence. This tripartite conceptualization of EBM is illustrated by the popular "three circle" model[7–9] as presented in Figure 1 of Appendix I. Despite the significance of empirical research as a core tenet of the EBM process, its then proponents claimed that clinical experience and patient preferences are indispensable to CDM, and can, at times, override research evidence[8(pA15)]. However, a frequently raised critique of this initial model (and future iterations) was the lack of explicit guidance on how to integrate the three components (expertise, research evidence and patient preferences)[10,11] and ultimately, put this model into practice[12].

In the early 2000s, the definition of EBM was further defined as: the "integration of best research evidence with clinical expertise and patient values" [13(pA11)]. A renewed model of evidence-based decisions was put forward which reflected certain changes to the original three circle model[13] (Figure 2 of Appendix I). Specifically, (1) the circle previously called *clinical expertise* was changed to *clinical state and circumstances* which reflected an acknowledgment of the uniqueness of each clinical encounter and the influence of context on EBM; (2) the oval labeled *clinical expertise* was drawn over the overlapping area of the three circles to depict its penultimate influence in CDM; and (3) the *patient preferences* circle was renamed *patient preferences and actions* suggesting that "patients' actions may differ from both their preferences and their clinician's advice" [13(pA12)]. This model accentuated the central role of clinical expertise and acknowledged the influence of clinical circumstances on CDM.

Following the publication of the 2005 *Sicily statement*, EBM proponents began to explicitly recognize that CDM occurs "within the context of available resources"[14(p1)]. This publication also marked the transition from EBM to evidence-based practice[a] (EBP) to be more inclusive to other professions such as nursing, rehabilitation and social work who were rapidly prescribing to the movement[14]. Indeed, several authors began to acknowledge the importance of context in CDM and incorporated contextual influences into their newer conceptualizations of EBP. For instance, the model for evidence-based social work by Regehr, Stern, and Shlonsky[15] placed the 2002 version of the EBP model within a broader contextual frame including the political, economic, and socio-historical context. In nursing, DiCenso, Ciliska and Guyatt added

[a] The term EBP will be used hereafter in this dissertation

a fourth circle to the 2002 EBP model labeled *healthcare resources*[16(p5)]. Similarly, Bannigan and Moores[17] presented a four-circle model of evidence-based decision making in occupational therapy consisting of evidence from research, clinical judgment, patient values and resources. Grounded in an ecological framework, Satterfield et al.[18] proposed the newest Transdisciplinary Model of EBP which, like others, recognizes the impact of the organizational context on CDM.

Despite the evolution in models of EBP, reflected in the addition of new components and in the subsequent iterations of the EBP model, many scholars continued to object to the absence of guidance on how a clinician can actually integrate the components (the "circles") in CDM[2,10,11,19]. Critics also argued that these models only provide elusive descriptions of EBP and lamented the absence of rationale for why the circles overlap only partially[20].

The doing of EBP

EBP is commonly operationalized as a five-step process represented by the five A's include: (1) *asking* a clinical research question; (2) *acquiring* the best available research evidence; (3) critically *appraising* the evidence for validity, clinical relevance, and applicability; (4) *applying* the results in practice; and (5) *assessing* the outcomes[14,18,21,22]. Curriculum developers in the health sciences were strongly encouraged to design courses for future healthcare professionals that would support the development of competencies (knowledge, skills, attitudes) grounded in the five-step model of EBP. In parallel, there was an expectation that clinicians carry out this stepwise process in practice[14]. *The Users' Guide to the Medical Literature* was widely disseminated across medical schools and reinforced the importance of this five-step process[23].

The five-step process became the first operationalization of EBP which served to clarify expectations for clinicians and to provide educators with a scaffolding of a complex process. Although the five-step process seemed broadly appealing, criticism abound: (1) it was criticized for being unrealistic in practice; (2) it was said to overemphasize critical appraisal at the expense of the last steps, apply and assess, which several argued were underdeveloped and lacked explicit guidance; (3) it undervalued clinicians' expertise, forcing them to adhere to a heuristic approach to CDM [24–28]. As Thomas, Chin-Yee and Mercuri[12] recently highlighted, descriptions of the five-step process have focused on where to find and how to critically appraise evidence while remaining elusive in the actual application of research findings to practice, a step that contains much uncertainty and contextual influence. Further, there is limited evidence that healthcare professionals actually enact the full five-step EBP process[29–34] . Specifically, rehabilitation clinicians have stated that although they believe in the importance and value of EBP, they do not believe it is their responsibility to undertake all five steps, specifically searching the literature and critically appraising articles[33,34]. Clinicians do, however, believe it is their role to determine the relevance of research evidence for the patient[33]. For example, findings from two survey studies in the last four years have suggested that critical appraisal was the least frequently used EBP step across healthcare professions[30] and that the item "engaging in all steps of the EBP process" ranked lowest in self-reported engagement in EBP among OTs and PTs[31]. Finally, prominent members of the EBP movement have abandoned the expectation that all clinicians can or should undertake critical appraisal of empirical studies[35,36]. Despite the limited evidence that the five steps are enacted in practice, the full EBP stepwise process continues to be promoted in guidelines for education and practice[37,38].

Main critiques on the early conceptualizations of EBP

Since the rise of the movement in the early 1990s, EBP has been met with widespread scrutiny. Djulbegovic et al.[39] highlight three main areas of criticism that the EBP movement has faced. The first criticism relates to what some have called the "dramatic irony" of EBP, itself not being evidence based as it fails to empirically demonstrate how EBP results in superior care and improved patient outcomes [24,40–42]. The second criticism relates to the lack of explicit guidance on how to integrate the various components in the EBP process when making clinical decisions [2,10,11,19] . The third criticism relates to the reductionist approach to clinical practice due to an over reliance on population-based evidence for the care of individual patients [41,43,44]. This third critique, which relates to the epistemological underpinnings of EBP, has been extensively discussed. For instance, what counts as "evidence" has been a central and contentious concept surrounding EBP. Early EBP proponents strongly emphasized observable empirical data derived from robust methods such as those from randomized controlled trials (RCTs) as the best research evidence[6]. However, this stance was heavily criticized for its assumption that relying on population-based findings invariably results in the best care for individual patients[11,27,43–46]. Early EBP proponents encouraged strict reliance on the evidence hierarchy, illustrated in the form of a pyramid, as a heuristic tool to judge the quality of evidence based on the methods used to generate it[23]. Meta-analyses, systematic reviews and RCTs constituted the top levels of the hierarchy as they were purported to reflect higher methodological rigor and minimize bias between independent and dependent variables compared to other types of study designs[7,47].

Defining optimal ways to produce knowledge and prioritizing a way of knowing about the world speaks to the epistemological positions of EBP[48]. Rooted in claims of clinical epidemiology and traced back to the 1970's when RCTs were promoted as unbiased evaluations of treatments, high quality evidence is equated with the application of standardized methods which establish linear causality, assure strong internal validity and best predict future outcomes[39,49]. As such, the epistemological principles inherent in the evidence hierarchy reflect positivist and empiricist views of inferences, judgment, and methods[39,48–50] . From this perspective, higher order evidence is more likely to lead researchers to an objective truth and should be prioritized in making decisions about the care of patients[51–53]. Scholars argued that such views of EBP are overly reductionist and rigid, in that they assume a view of evidence that ignores context and the individual biological and social differences between patients[54–58]. While RCTs results can be valuable (they have in fact greatly advanced our understanding of interventions and treatments), the application of findings to practice which is rife with complexity, heterogeneity and uncertainty must be deliberately considered to avoid prescribing population-based results to individuals[44,48].

Shift towards a contemporary conceptualization of evidence and EBP

As it became increasingly recognized that one's understanding of what constitutes a legitimate source of evidence was a matter of epistemological perspective, the traditional and narrow concept of evidence began to evolve[26]. EBP pioneers published position statements acknowledging the limitations of inferences derived from RCTs whilst recognizing the value of evidence derived from study designs such as case studies, observational and qualitative studies[59].

Throughout the years, EBP evolved from a purely positivist, empiricist approach to CDM to a more pluralist stance acknowledging the complexity and uncertainty inherent to clinical practice and the contextual nature of evidence[57,60]. As such, EBP is increasingly being conceptualized as a socially constructed, dynamic process[61,62]. From a social constructivist perspective, "research becomes integrated with previously held knowledge, and humans build on and create knowledge through their interactions with each other"[63(p2)]. Clinicians do not passively receive research-based information, they actively make sense of the findings, interpret the information, and transform it when making clinical decisions[48,64–67]. Further, the clinician participates in knowledge exchange with the patient, the patient's support system and the interdisciplinary team which plays a role in developing knowledge[61,64]. Situating EBP within a social constructivist paradigm recognizes the importance of different sources of knowledge used in CDM including tacit knowledge, expert opinion, physiological or basic scientific knowledge, and patient preferences and values[2,24,44,68–70]. A social constructivist perspective also acknowledges the co-construction of knowledge through social interactions[64,65,71]. While scholars claimed that early conceptualizations of EBP risked damaging the autonomy and integrity of clinicians[72,73], this modern understanding of EBP honors autonomy and agency and recognizes expertise as vital in the consolidation and use of knowledge in CDM[43,66].

As described in a three-part essay published in the *Journal of Evaluation in Clinical Practice*, Mugerauer claims that "we appear to be at the beginning of a post-EB[P] period that is combining new approaches with the return to core values" [74(p593)]. Mugerauer argues for "a modified original, moderate position in which clinical decision making is person-centered, recognizing and integrating multiple modes of evidence and knowledge that have been marginalized: professional experience, illness narratives, and individual patients' values and preferences."[74(p593)] In alignment with Mugereauer's stance, the modern view of EBP has become more epistemically balanced, shifting from the dichotomy of *objective* versus *subjective*[75] to a recognition of the multiple sources of research evidence and types of knowledge inherent to the EBP process[76]. While the EBP debate has moved forward and recognizes the complexity of CDM, one cannot understate the central axiom of EBP - the integration of best available research evidence into the CDM process still serves a critical role[2,48]. Rather than firmly relying on heuristic tools like an evidence hierarchy or the five-step process, the cornerstone of EBP should be the continued questioning of the relationship between research evidence and practice[77(p222)].

EBP in rehabilitation

A professional expectation for OTs and PTs

The World Health Organisation defines rehabilitation as "a set of interventions designed to optimize functioning and reduce disability in individuals with health conditions in interaction with their environment"[78]. Currently, approximately 2.4 billion people worldwide have a health condition requiring rehabilitative care; this number is expected to dramatically increase with the doubling of individuals over 60 years of age by 2050 and the rise of chronic conditions such as diabetes, stroke, and cancer[78]. There is an ongoing need for effective rehabilitation services[78].

EBP can optimize quality of care through the provision and application of research evidence on the effectiveness of rehabilitation approaches and can contribute to reducing ineffective and potentially harmful practices[79,80]. Integrating best available evidence into CDM also responds to expectations from consumers, funding agencies and regulatory bodies for high-quality, cost-effective care[79,81–83]. EBP can also contribute to professional accountability and credibility as integrating research evidence into practice can serve to justify clinical decisions[14,26,84–86]. Clinicians have also reported that EBP fosters professional empowerment and job satisfaction as it can prompt self-directed life-long learning[61,87–89]. On the reverse, a lack of engagement in the EBP process may lead to missed opportunities to benefit patients and society and it may be considered unethical to withhold research-based evidence about existing and available treatments[90].

For the expected benefits on provision of services and health outcomes to materialise, national professional associations, regulators and associations of university programs advocate for an evidence-based approach to CDM. Canadian OTs and PTs are expected to integrate research evidence actively and sustainably into their CDM process about the assessment and/or treatment of patients, as articulated in both professional competency profiles[81,91].

The integration of research evidence into decision-making is lagging in rehabilitation

Despite the anticipated benefits and the growing emphasis on EBP, the integration of research evidence into CDM for OTs and PTs remains a major challenge[30,31,34,92–98]. Two 2020 studies demonstrated that two-thirds of recently graduated Canadian OTs and PTs report engaging in EBP[95], and that OTs in Quebec self-rated, on average, their competency level with the scholar role (the one that incorporates EBP) to be 30%, placing this role sixth among the seven core competencies[96]. A 2014 systematic review of 32 studies on OTs' engagement with EBP indicated a clear and consistent trend in low utilization of research evidence[93]. In a 2018 cross-sectional survey of over 600 healthcare professionals involved in pain management, the EBP behaviors scores for OTs and PTs were amongst the lowest and were considerably lower than their scores on the knowledge or attitudes subscales[97].

In other countries, studies have found comparable trends demonstrating the challenge in integrating research evidence into the CDM process. Two 2020 surveys demonstrated that most American OTs and PTs engaged with the EBP process only "some of the time"[31] and that only 60% of PTs in Saudi Arabia reported adopting EBP on a regular basis[98]. In Sweden, a 2014 survey found that 53% of PTs reported using guidelines sometimes, infrequently, or never[92] and a 2018 survey of four healthcare professional groups reported that OTs scored lowest in self-reported use of EBP[30]. Earlier studies in Oceanian countries demonstrate that only 56% of OTs referenced the research literature to guide their CDM in the past two months[99] and that OTs self-reported low to moderate scores in practices related to the utilization of research[100].

Together, these studies suggest that consulting and integrating research evidence into the CDM process remain important challenges. Despite aggregated efforts to improve engagement with the EBP process, there is a need to support rehabilitation clinicians' use of research evidence to inform their CDM process and to develop a culture where integration of research evidence is expected and fostered[83].

EBP determinants in rehabilitation

The last three decades have been marked with important advancements on the individual and organizational factors that impede or facilitate the enactment of EBP by rehabilitation professionals[84,95,101–105]. Particularly, there have been several studies that have examined the association between a clinician's knowledge, skills, attitudes, and self-efficacy on EBP behaviors[34,95,97,99,106].

Knowledge is defined as the "retention of facts and concepts about EBP" such as defining EBP concepts, listing basic principles of EBP, and describing levels of evidence"[37(pp4-5)]. A recent survey of new graduates from the 29 occupational and physical therapy programs in Canada found overall good levels of knowledge about statistical terms and methodological concepts[95]. However, no association was found between the *knowledge* scale and the *use of research evidence* or *activities related to EBP* scales which suggests that while knowledge levels on statistical concepts are adequate, they have little impact on actual self-reported use of EBP in practice[95]. This may relate to the low enactment of critical appraisal steps by rehabilitation clinicians[29–31,34,93] and the existence of clinical practice guidelines and decision-support tools[36]. While knowledge of statistical terms may not be required to be an evidence-based clinician, knowledge on the existence of specific evidence-based interventions is critical[103]. A study which surveyed OTs in Quebec found that only 53% agreed or strongly agreed that they were aware of the evidence-based practices in their area of practice[96] which suggests that clinicians must be supported in keeping up to date with knowledge on best practices.

Skills is defined as "the application of knowledge, ideally in a practical setting" such as conducting a search or using a critical appraisal tool[37(pp4-5)]. Rehabilitation clinicians have reported low to moderate levels of skills in critically appraising the research literature and interpreting statistical analyses and results[101,107–109] . Similar levels were found by Thomas et al.[95] whereby a third of the sample of recently graduated OTs and PTs reported difficulty in critical appraisal of study methods. Although OTs and PTs may report low to moderate levels of skills in critical appraisal, critical appraisal is no longer considered a necessary requirement to engage in EBP given the rise of pre-appraised knowledge products such as clinical practice guidelines and decision support tools[36].

Attitudes is defined as "the values ascribed [...] to the importance and usefulness of EBP to inform [CDM]" [37(pp4-5)]. While attitudes have been studied as precursors to behavior change[110,111], findings demonstrating an association between attitudes towards EBP and actual integration of research evidence is inconclusive. Several studies report that having positive attitudes is an important facilitator to enactment of the EBP process[95,112–114]. However, other studies suggest that despite positive attitudes of OTs and PTs towards EBP, these favorable attitudes do not translate into EBP behaviors; rehabilitation clinicians continue to be challenged with the integration of research evidence[34,84,92,97,99,107,115]. Nevertheless, rehabilitation clinicians generally hold highly favorable attitudes towards EBP[31,34,84,92,95,99,107,113–116].

Self-efficacy is defined as "people's judgments regarding their ability to perform a certain activity"[37(pp4-5)]. Self-efficacy is considered to be an important psychological construct for understanding EBP behavior in healthcare professionals[117–119]. Higher levels of self-efficacy in enacting EBP are associated with higher self-reported use of research evidence in CDM[95,120]. Data from cross-sectional surveys showed that only about half of Swedish OTs felt confident in finding and using evidence[116] and that OTs in New Zealand reported low to moderate levels of confidence in research-related skills[99].

It has also been suggested that specific activities in which clinicians participate have an impact on the use of EBP[61,121]. For instance, OTs who frequently mentor trainees self-rate higher on perceived competence as a scholarly practitioner[96]. Other activities associated with higher engagement with EBP include participation in journal clubs, conferences or research projects, informally discussing research findings with colleagues and being subscribed to article alerts[86,122–125].

It is also well-recognized that the organizational context and culture impact the implementation of research evidence[90,100,104,105,107,126–128]. Recently, Paci et al. conducted a systematic review and meta-analysis of 29 articles on barriers to EBP in physical therapy across a range of practice settings and found that the principal barrier was lack of protected time for activities related to EBP such as reading the literature followed by the lack of access to resources such as research databases[101]. In occupational therapy, the main organizational-level barriers to EBP include the lack of protected time to access, read and implement best practices, inadequate equipment to implement evidence-based interventions, and low human and financial resources[93,103].

These interacting individual and organizational influences on the EBP process highlight the complexities of implementing research evidence into practice and support the characterization of EBP as a multidimensional process. It has been hypothesized that the complex and multifactorial nature of the EBP process may be responsible for the variable use of EBP among rehabilitation clinicians[34]. It is thus paramount to consider the multidimensional nature of EBP in future studies.

Measurement of EBP

In parallel to the steadfast emphasis on promoting EBP, there has been growing interest in its measurement. Robust measurement practices are essential to determine if clinicians are engaging in the EBP process and to identify the factors related to EBP that should be improved or the strong areas that must be maintained[129–131] to better support clinicians and ultimately, improve quality of care[129,130].

Current measurement of EBP

In the past two decades, EBP measurement methods have rapidly developed to cover a wide range of purposes, constructs, and scale constructions[37]. Approaches to measuring EBP can be categorized into (1) external assessment or observation, (2) achievement tests of knowledge and skills, and (3) self-report questionnaires [132].

External assessment or observation

External assessments or observations focus on recording the level of EBP behavior that clinicians enact during a determined period of time. This approach, which can include direct observation of a clinician or the use of a proxy measure such as chart audits or video-stimulated recall, are often labeled *objective* as they do not rely on retrospective self-report[132]. In a systematic review of 115 articles across the healthcare professions, Shaneyfelt et al. demonstrated that these formats are rarely employed with only three external or observational assessments of 104 identified methods[132]. Hrisos et al. suggest that a possible reason for the low use of external observations for measuring EBP is that "it is often not feasible or ethical to measure behavior through direct observation"[133(p1)]. Indeed, observation can be intrusive to the clinician and the patient, can be prone to social desirability bias as someone is being watched and, in many cases can be resource-intensive (i.e., both timely and costly)[133].

Chart audits are also considered external assessments by proxy; however, this approach is said to introduce large amounts of variability in data collection and analysis[134] and entail long hours of analysis[135]. Additionally, chart reviews may miss important pieces of information as clinicians underreport their behaviors[135,136] and some aspects of clinical acumen, especially when related to counseling and educational interventions, cannot be explicitly and rapidly recorded[133].

Achievement tests of knowledge and skills

Achievement or cognitive tests measure the extent to which an individual has met pre-established standards with regards to knowledge and/or skills. These can include short answer, essay, or multiple-choice questions and involve standardized procedures providing consistency and centralized scoring[137–139]. Such tests would require the clinician to attain a minimal level of knowledge or skill level to be considered competent to practice. These tests may be time-consuming to develop as the minimum standards and criteria must be determined[137,140]. Tilson et al. suggest using cognitive testing to measure EBP knowledge because self-report approaches measure the knowledge-level as perceived by the clinician (not the knowledge itself)[37]. A systematic review on EBP measures in occupational therapy by Glegg and Holsti identified only one achievement test used with OTs compared to 14 self-reported questionnaires[141]. However, as mentioned earlier, having the prerequisite knowledge and skills does not necessarily translate to EBP enactment in practice[95,97].

Self-report questionnaires

Self-report questionnaires are undoubtedly the most frequently used approach to measuring EBP across the healthcare professions[129,130,132,141,142]. While the self-report format has been criticized for being prone to social desirability bias[132] (like the external observation approach), there are several benefits namely, the measurement of constructs that are not observable such as self-efficacy, their pragmatic nature (i.e., ease and low cost of administration) and the overall acceptability by respondents[141,143].

In occupational therapy and physical therapy, there exist at least 34 and 24 measures, respectively, to assess EBP domains such as knowledge, skills, attitudes, self-efficacy, or behaviors[129,142]. Some questionnaires only focus on a single EBP domain such as self-efficacy (Evidence-based Practice Confidence Scale by Salbach et al.[120]) or attitudes (Questions on EBP questionnaire by Stevenson et al.[144]). Since it is desirable to assess more than one domain due to the multidimensional nature of EBP[129,130], several unidimensional questionnaires must be administered to gain a comprehensive understanding of a clinician's EBP engagement. This approach is problematic for two reasons: first, employing multiple questionnaires is burdensome for clinicians and labor intensive for researchers, and second, some of these measures may overlap in terms of what they measure, a form of redundancy that can lead to non-completion and false reliability[131,145].

There also exist self-report questionnaires that combine multiple domains in various permutations. A systematic review by Buchanan et al. identified eight recommended EBP measures for use in occupational therapy; these measures comprise different permutations of domains including EBP knowledge (n=5), skills in EBP (n=1), attitudes to EBP (n=5), and EBP behavior (n=8)[142]. Descriptive information on these eight EBP measures can be found in Appendix II. Of note, none of these eight measures comprehensively include all domains of knowledge, attitudes, and skills. A similar systematic review of EBP measures in physical therapy identified 24 self-report questionnaires but concluded that all had important shortcomings regarding the inclusion of EBP domains and psychometric evidence[129]. Despite the availability of questionnaires, Fernandez-Domínguez et al. also noted that none of the questionnaires measured the organizational or contextual factors that may influence the enactment of EBP[129].

Since the publication of these two systematic reviews in 2016 (occupational therapy) and 2014 (physical therapy), two new measures of EBP have been published including the *Health Sciences Evidence Based Questionnaire*[146] developed in 2017 and the questionnaire by Al Zoubi et al. developed in 2018[131]; these are also described in Appendix II. Unlike earlier questionnaires identified in the systematic reviews, these newer questionnaires include items assessing organizational level factors in addition to items focused on the individual clinician, suggesting a trend towards more comprehensive measures and an acknowledgment of EBP as contextually dependent.

Challenges with current measurement of EBP

Despite the availability of multiple approaches to measuring EBP, limitations with existing measures jeopardize the quality of the data that they produce and may lead to misinterpretation of results. Six key measurement issues, related to the implementation of best practices, have been identified by Martinez et al. in 2014 including (1) the various and overlapping theories, models and frameworks which lead to construct confusion; (2) the inconsistent conduct and reporting of appropriate psychometric tests; (3) the practice of developing one-time use home-grown instruments which limits the capacity for cross-study comparison; (4) the complexity of choosing the most appropriate evaluation method for the research question; (5) the need for practical instruments (i.e., low costs, item count lower than 10) for use in resource-demanding settings; and (6) the need for decision-making supports in choosing the appropriate instrument[130].

Another important challenge with the majority of EBP questionnaires in rehabilitation relates to the inappropriate treatment of ordinal data such as data derived from Likert-type scales. Likert-type scales are popular because it is more intuitive and straight-forward to respond to textual responses over numerical rating scales[147–149]. Often, numbers are attributed to the categories of the Likert-type scale and authors sum or average across items[150,151]. However, the allocation of numbers is not indicative of any mathematically valid magnitude of difference between categories[152]. Thus, performing arithmetic operations (e.g., mean, standard deviation) and parametric inferential statistics (e.g., ANOVA) on ordinal data is a flawed approach to analysis and can generate misleading conclusions[150,151].

In developing a comprehensive measure of EBP in rehabilitation, Al Zoubi et al.[131] were the first, in the context of EBP in rehabilitation, to use Rasch analysis to sum across items and generate a mathematically accurate total score per construct. Rasch Measurement Theory (RMT) is a modern measurement approach for the modeling of categorical data[153]. RMT allows for the mapping of unidimensional constructs onto a single, linear scale from the easiest to the most difficult item using a logit transformation[154,155]. In parallel, RMT also estimates the spread of respondents on a continuum based on their ability levels (least able to most able). Rasch analysis can assess item fit to the Rasch model, whether items span the entire construct, the extent to which items have response choices that are appropriately ordered and whether items perform differently between groups of people[153,156,157]. Results from Rasch analysis allow researchers to reduce the number of items and rewrite or exclude problematic items (i.e., items prone to bias). This allows for a more rapid and less burdensome test experience. Furthermore, calibrating items onto the same linear scale allows for a mathematically valid approach to summing scores by demonstrating the interval-like properties of items.

Important progress has been made in the development of EBP measures, notably regarding construct comprehensiveness of EBP questionnaires (i.e., the inclusion of individual and organizational domains) and the use of modern measurement theory to reduce the number of problematic items and establish a mathematical basis for scoring. However, existing multidimensional EBP self-report questionnaires can be considered as *profiles* whereby results are summarized and interpreted by sub-scale, where each sub-scale represents an EBP domain[158]. With this profile-type approach, the relative importance of one domain in relation to others is unaccounted for resulting in a fragmented interpretation of EBP. As such, one cannot combine the domains for a broad interpretation of EBP (i.e., a total score) that would allow for trade-offs to be made between domains. For example, if a sample demonstrates high self-efficacy but low levels of knowledge and skills, it is difficult to arrive at a conclusion on their combined interpretation. Yet, by virtue of the multidimensional nature of EBP, individual and organizational factors synergistically affect whether a clinician will engage with the EBP process. This interdependency between factors must be considered when measuring EBP.

A similar measurement complexity exists in the field of health-related quality of life (HRQoL). An important body of research has refined the methods used to measure HRQoL, a complex multidimensional construct, and developed preference-based measures (PBMs) such as the Health Utilities Index (also called the HUI), the EuroQol-5D (also called the EQ-5D) and the Short Form-6D (also called the SF-6D)[159,160]. These PBMs consider the key domains influencing HRQoL and their relative importance in generating a single score representative of the overall health state based on what is most important to the patient group[159,161–164]. PBMs can be generic or condition specific (e.g., multiple sclerosis, stroke, and obesity)[165–167] and the most popular PBMs are brief, typically consisting of one item per domain (called a dimension). The main advantage of PBMs is their ability to generate an overall single score that balances gains in one domain versus losses in another domain[165–167].

To determine the relative weights of each dimension contained within a PBM, individuals are asked to provide input on the relative importance of certain domains compared to others (i.e., their preferences) using one or more weight elicitation techniques[168–170]. Many techniques have been reported in the literature and can be classified into two broad approaches: choice-based exercises and direct elicitation of numerical values[171].

The first approach was developed in marketing research to predict buyer behavior and aims to determine how changes in specific product attributes can influence choice[171,172]. This approach can involve ranking, rating, or choice designs (e.g., conjoint analysis) where respondents are asked to consider and evaluate features or products relative to others. The second approach consists of methods using direct elicitation of numerical values (e.g., rating scale or willingness-to-pay). This approach aims to estimate the demand for (or value of) a specific attribute within a product[171]. The selection of a method depends on the fit with the study perspective. Based on the statistical analysis associated with the selected weight elicitation technique, weights can be predicted for all levels (response options) within a dimension (i.e., dimension-levels). These estimated weights can then be integrated into a scoring algorithm that enables the generation of a total score.

To summarize, while many EBP questionnaires exist, there are important challenges in how EBP is currently measured which may affect the accuracy of inferences. First, many questionnaires do not cover the most salient individual *and* organizational EBP domains resulting in a need to administer multiple questionnaires to gain a comprehensive portrait of EBP which can increase respondent burden and chances for dropouts and item redundancy. Second, the analysis of ordinal data from EBP questionnaires is often inappropriately treated as interval-level data resulting in misleading interpretations. Third, the interpretation of EBP is fragmented due to "profile-type" measures which summarize results of domains independently. Finally, the relative importance of the domains that influence EBP and how these can be quantitatively captured in a measure are unknown. These measurement challenges provide the central impetus for this doctoral dissertation.

Positionality

In this thesis, *EBP* is defined as a complex and multidimensional approach to CDM that combines the best available research evidence, clinical expertise, and patients' preferences, within the context of available resources[14,26,173]. This conceptualization of EBP does not advocate for a mechanistic and prescriptive application of research evidence but a conscientious and judicious consideration of best available evidence into the CDM process[7]. I view the EBP process as including and being dependent upon, in no order, research-based knowledge, tacit knowledge, expert opinion, and patient preferences and values [2,24,44,68–70]. In this thesis, I define *research evidence* as the form of knowledge derived from academic research [49,68,174–176] and non-research-based warrants for CDM are termed *sources of knowledge*.

"Measurement itself, like scientific research in general, is an epistemic activity- that is, an attempt to gain knowledge about the world."[177(p756)] The science and practice of measurement are closely associated with a post-positivist theoretical perspective which posits that there is a true object or social reality "out there" that has enough stability and patterning to be known, albeit in an imperfect manner[178,179]. As stated by Creswell, researchers within a post-positivist perspective examine individuals' behavior through numeric measurements[180]. As such, this thesis, given its focus on measurement, is situated within a post-positivist research paradigm while recognizing EBP as a socially constructed phenomenon, the central idea being that a clinician actively makes sense of and integrates research-based knowledge with previously held knowledge and applies this knowledge in socially mediated contexts[3,32,49,64,65,181].

Given the above, EBP cannot be directly observed in the same way as objects (e.g., length, mass, time, temperature, etc.), as it is a socially constructed, multidimensional, context-dependent, and dynamic process. EBP is a latent construct, a variable that cannot be observed but can be estimated and inferred by related variables[158]. Further, certain components of EBP such as tacit knowledge and expert opinion are immensurable[39,181]. In this view, it would be unjustified to claim that questionnaires measure EBP if one views EBP as including multiple sources of knowledge, yet fails to capture these sources of knowledge. Most of the existing literature claims to measure EBP but does not seek to capture these other sources of knowledge.

By expanding on previous measurement standards and an accepted body of knowledge, this doctoral research responds to two methodological commitments put forward by Chang, a philosopher of science and measurement expert. The first principle, “the principle of respect”, advises to “respect the prior [measurement] standard as far as it is plausible to do so” while the second, “the imperative of progress”, recommends that scientists continue to improve the epistemic virtues of their predecessors[182(p44)].

References

1. Pritchard MJ. Professional development. Making effective clinical decisions: a framework for nurse practitioners. *Br J Nurs*. 2006;15(3):128-130. doi:10.12968/bjon.2006.15.3.20507
2. Tonelli. Integrating evidence into clinical practice: an alternative to evidence-based approaches: Integrating evidence into clinical practice. *J Eval Clin Pract*. 2006;12(3):248-256. doi:10.1111/j.1365-2753.2004.00551.x
3. Loughlin M, Bluhm R, Buetow S, Borgerson K, Fuller J. Reasoning, evidence, and clinical decision-making: The great debate moves forward. *J Eval Clin Pract*. 2017;23(5):905-914. doi:10.1111/jep.12831
4. Djulbegovic B, Elqayam S. Many faces of rationality: Implications of the great rationality debate for clinical decision-making. *J Eval Clin Pract*. 2017;23(5):915-922. doi:10.1111/jep.12788
5. Guyatt GH. Evidence-based medicine. *ACP J Club*. Published online 1991:A-16.
6. EBM working group. Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama*. 1992;268(17):2420.
7. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn’t. *BMJ*. 1996;312(7023):71-72.
8. Haynes RB, Sackett DL, Gray JMA, Cook DJ, Guyatt GH. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. *Evid Based Med*. 1996;1(7):196.
9. Haynes RB, Sackett DL, Richardson WS, Rosenberg W, Langley GR. Evidence-based medicine: How to practice & teach EBM. *Can Med Assoc J*. 1997;157(6):788.
10. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *The Lancet*. 1995;345(8953):840-842.
11. Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. *The Lancet*. 1997;349(9045):126-128.
12. Thomas A, Chin-Yee B, Mercuri M. Thirty years of teaching evidence-based medicine: have we been getting it all wrong? *Adv Health Sci Educ*. 2022;27(1):263-276.

13. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *BMJ Evid-Based Med*. 2002;7(2):36-38.
14. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1
15. Regehr C, Stern S, Shlonsky A. Operationalizing evidence-based practice: The development of an institute for evidence-based social work. *Res Soc Work Pract*. 2007;17(3):408-416.
16. DiCenso A, Ciliska D, Guyatt G. Introduction to evidence-based nursing. *Evid-Based Nurs Guide Clin Pract*. Published online 2005:3-19.
17. Bannigan K, Moores A. A model of professional thinking: Integrating reflective practice and evidence based practice. *Can J Occup Ther*. 2009;76(5):342-350.
18. Satterfield JM, Spring B, Brownson RC, et al. Toward a Transdisciplinary Model of Evidence-Based Practice. *Milbank Q*. 2009;87(2):368-390. doi:10.1111/j.1468-0009.2009.00561.x
19. Upshur REG. The complex, the exhausted and the personal: reflections on the relationship between evidence-based medicine and casuistry. Commentary on Tonelli (2006), Integrating evidence into clinical practice: an alternative to evidence-based approaches. Journal of Evaluation in Clinical Practice 12, 248–256. *J Eval Clin Pract*. Published online 2006:8.
20. Charles C, Gafni A, Freeman E. The evidence-based medicine model of clinical practice: scientific teaching or belief-based preaching? *J Eval Clin Pract*. 2011;17(4):597-605. doi:10.1111/j.1365-2753.2010.01562.x
21. Cook DJ, Jaeschke R, Guyatt GH. Critical appraisal of therapeutic interventions in the intensive care unit: human monoclonal antibody treatment in sepsis. Journal Club of the Hamilton Regional Critical Care Group. *J Intensive Care Med*. 1992;7(6):275-282. doi:10.1177/088506669200700601
22. Sackett DL. Evidence-based medicine. In: *Seminars in Perinatology*. Vol 21. Elsevier; 1997:3-5.
23. Guyatt Gordon, Rennie Drummond, Evidence-Based Medicine Working Group., American Medical Association. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. AMA Press; 2002.

24. Norman GR. Examining the assumptions of evidence-based medicine: Examining the assumptions of EBM. *J Eval Clin Pract*. 1999;5(2):139-147. doi:10.1046/j.1365-2753.1999.00197.x
25. Barratt A. Evidence based medicine and shared decision making: the challenge of getting both evidence and preferences into health care. *Patient Educ Couns*. 2008;73(3):407-412.
26. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014
27. Vandenbroucke JP. Observational Research and Evidence-Based Medicine: What Should We Teach Young Physicians? *J Clin Epidemiol*. 1998;51(6):467-472.
28. McAlister FA, Graham I, Karr GW, Laupacis A. Evidence-based medicine and the practicing clinician. *J Gen Intern Med*. 1999;14(4):236-242.
29. Condon C, McGrane N, Mockler D, Stokes E. Ability of physiotherapists to undertake evidence-based practice steps: a scoping review. *Physiotherapy*. 2016;102(1):10-19.
30. Boström AM, Sommerfeld DK, Stenhols AW, Kiessling A. Capability beliefs on, and use of evidence-based practice among four health professional and student groups in geriatric care: A cross sectional study. *PLOS ONE*. 2018;13(2):e0192017. doi:10.1371/journal.pone.0192017
31. Wang Y, King K, Moran B, et al. Occupational and Physical Therapists' Perception of Evidence-Based Practice. *J Allied Health Wash*. 2019;48(2):119-126,126A-126C.
32. Gabbay J, May A le. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. *BMJ*. 2004;329(7473):1013. doi:10.1136/bmj.329.7473.1013
33. Salbach NM, Jaglal SB, Korner-Bitensky N, Rappolt S, Davis D. Practitioner and Organizational Barriers to Evidence-based Practice of Physical Therapists for People With Stroke. *Phys Ther*. 2007;87(10):1284-1303. doi:10.2522/ptj.20070040
34. Scurlock-Evans L, Upton P, Upton D. Evidence-Based Practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014;100(3):208-219. doi:10.1016/j.physio.2014.03.001
35. Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence based care. *BMJ*. 2000;320(7240):954-955.

36. Tikkinen KA, Guyatt GH. Understanding of research results, evidence summaries and their applicability—not critical appraisal—are core skills of medical curriculum. *BMJ Evid-Based Med*. Published online 2021.
37. Tilson JK, Kaplan SL, Harris JL, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Educ*. 2011;11(1):1-10. doi:10.1186/1472-6920-11-78
38. Albarqouni L, Hoffmann T, Straus S, et al. Core competencies in evidence-based practice for health professionals: consensus statement based on a systematic review and Delphi survey. *JAMA Netw Open*. 2018;1(2):e180281-e180281.
39. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic Inquiries in Evidence-Based Medicine. *Cancer Control*. 2009;16(2):158-168. doi:10.1177/107327480901600208
40. Couto JS. Evidence-based medicine: a Kuhnian perspective of a transvestite non-theory. *J Eval Clin Pract*. 1998;4(4):267-275. doi:10.1111/j.1365-2753.1998.tb00085.x
41. Miles A, Bentley P, Polychronis A, Grey J. Evidence-based medicine: why all the fuss? This is why. *J Eval Clin Pract*. 1997;3(2):83-86. doi:10.1046/j.1365-2753.1997.00103.x
42. Miles A, Bentley P, Polychronis A, Grey J, Price N. Recent progress in health services research: on the need for evidence-based debate. *J Eval Clin Pract*. 1998;4(4):257-265. doi:10.1111/j.1365-2753.1998.tb00084.x
43. Greenhalgh T, Howick J, Maskrey N, for the Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348(jun13 4):g3725-g3725. doi:10.1136/bmj.g3725
44. Tonelli M. The philosophical limits of evidence-based medicine. Published online 1998.
45. Rogers WA. Is there a tension between doctors' duty of care and evidence-based medicine? *Health Care Anal*. 2002;10(3):277-287.
46. Upshur RE, Tracy CS. Legitimacy, authority, and hierarchy: Critical challenges for evidence-based medicine. *Brief Treat Crisis Interv*. 2004;4(3).
47. Sackett DL. Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest*. 1986;89(2, Supplement):2S-3S. doi:10.1378/chest.89.2_Supplement.2S
48. Kelly MP. The need for a rationalist turn in evidence-based medicine. *J Eval Clin Pract*. 2018;24(5):1158-1165. doi:10.1111/jep.12974

49. Ashcroft RE. Current epistemological problems in evidence based medicine. *J Med Ethics*. 2004;30(2):131-135. doi:10.1136/jme.2003.007039

50. Greenhalgh, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med*. 2018;16(1):95, s12916-018-1089-4. doi:10.1186/s12916-018-1089-4

51. Sackett DL, Haynes RB, Tugwell P, Guyatt GH. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Little, Brown Boston; 1985.

52. Davidoff F. Evidence-Based Medicine: Why All The Fuss? *Ann Intern Med*. 1995;122(9):727. doi:10.7326/0003-4819-122-9-199505010-00012

53. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Rethinking bias and truth in evidence-based health care. *J Eval Clin Pract*. 2018;24(5):930-938. doi:10.1111/jep.13010

54. Grahame-Smith D. Evidence based medicine: Socratic dissent. *BMJ*. 1995;310(6987):1126.

55. The Lancet. Evidence-based medicine, in its place. *The Lancet*. 1995;346(8978):785. doi:10.1016/S0140-6736(95)91610-5

56. Buetow S, Kenealy T. Evidence-based medicine: the need for a new definition: Examining the meaning of EBM. *J Eval Clin Pract*. 2000;6(2):85-92. doi:10.1046/j.1365-2753.2000.00237.x

57. Feinstein AR, Horwitz RI. Problems in the “evidence” of “evidence-based medicine.” *Am J Med*. 1997;103(6):529-535.

58. Howick JH. *The Philosophy of Evidence-Based Medicine*. John Wiley & Sons; 2011.

59. Sackett DL, Wennberg JE. *Choosing the Best Research Design for Each Question: It’s Time to Stop Squabbling over the “Best” Methods*. British Medical Journal Publishing Group; 1997.

60. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM. *J Eval Clin Pract*. 2017;23(5):964-970. doi:10.1111/jep.12752

61. Hallé MC, Mylopoulos M, Rochette A, et al. Attributes of evidence-based occupational therapists in stroke rehabilitation. *Can J Occup Ther*. 2018;85(5):351-364. doi:10.1177/0008417418802600

62. Reagon C, Bellin W, Boniface G. Challenging the Dominant Voice: The Multiple Evidence Sources of Occupational Therapy. *Br J Occup Ther*. 2010;73(6):284-286. doi:10.4276/030802210X12759925469069
63. Kothari AR, Bickford JJ, Edwards N, Dobbins MJ, Meyer M. Uncovering Tacit Knowledge: A Pilot Study to Broaden the Concept of Knowledge in Knowledge Translation. *BMC Health Serv Res*. 2011;11(1):1-10. doi:10.1186/1472-6963-11-198
64. Thomas A, Saroyan A, Dauphinee WD. Evidence-based practice: a review of theoretical assumptions and effectiveness of teaching and assessment interventions in health professions. *Adv Health Sci Educ*. 2011;16(2):253-276. doi:10.1007/s10459-010-9251-6
65. Thomas A, Menon A, Boruff J, Rodriguez AM, Ahmed S. Applications of social constructivist learning theories in knowledge translation for healthcare professionals: a scoping review. *Implement Sci*. 2014;9(1):1-20.
66. Wieten S. Expertise in evidence-based medicine: a tale of three models. *Philos Ethics Humanit Med*. 2018;13(1):2. doi:10.1186/s13010-018-0055-2
67. Rycroft-Malone J. Promoting action on research implementation in health services (PARIHS). *Models Framew Implement Evid-Based Pract Link Evid Action*. 2010;109:135.
68. Gupta M. Reconsidering rationality and ethics in the evidence-based medicine debate: a reply to commentators. *J Eval Clin Pract*. 2004;10(2):143-146. doi:10.1111/j.1365-2753.2004.00500.x
69. Solomon M. Just a paradigm: evidence-based medicine in epistemological context. *Eur J Philos Sci*. 2011;1(3):451-466. doi:10.1007/s13194-011-0034-6
70. Estabrooks CA. Will evidence-based nursing practice make practice perfect? *Can J Nurs Res Arch*. Published online 1998:15-36.
71. Shay S. Beyond social constructivist perspectives on assessment: the centring of knowledge. *Teach High Educ*. 2008;13(5):595-605. doi:10.1080/13562510802334970
72. Kinsella EA, Whiteford GE. Knowledge generation and utilisation in occupational therapy: Towards epistemic reflexivity. *Aust Occup Ther J*. 2009;56(4):249-258.
73. Miettinen OS, Flegel KM. Professionalism in medicine. *J Eval Clin Pract*. 2003;9(3):353-356.

74. Mugerauer R. Professional judgement in clinical practice (part 1): Recovering original, moderate evidence-based health care. *J Eval Clin Pract*. 2021;27(3):592-602. doi:10.1111/jep.13513
75. Loughlin M, Upshur RE, Goldenberg MJ, Bluhm R, Borgerson K. Philosophy, ethics, medicine and health care: the urgent need for critical practice. Published online 2010.
76. Shlonsky A, Mildon R. Methodological pluralism in the age of evidence-informed practice and policy. *Scand J Public Health*. 2014;42(13_suppl):18-27.
77. Dent J, Harden RM, Hunt D. *A Practical Guide for Medical Teachers, E-Book*. Elsevier Health Sciences; 2021.
78. World Health Organization. Rehabilitation. Published November 2021. https://www.who.int/news-room/fact-sheets/detail/rehabilitation
79. Emparanza JI, Cabello JB, Burls AJE. Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *J Eval Clin Pract*. 2015;21(6):1059-1065. doi:10.1111/jep.12460
80. Leland NE, Crum K, Phipps S, Roberts P, Gage B. Advancing the Value and Quality of Occupational Therapy in Health Service Delivery. *Am J Occup Ther*. 2015;69(1):6901090010p1-6901090010p7. doi:10.5014/ajot.2015.691001
81. ACOTRO, ACOTUP, CAOT. Competencies for Occupational Therapists in Canada/Référentiel de compétences pour les ergothérapeutes au Canada. Published online 2021. Accessed July 12, 2022. https://acotro-acore.org/sites/default/files/uploads/ot_competency_document_en_hires.pdf
82. Murphy S, Lin S. Perspectives. Keeping occupational therapy alive in the 21st century: The role of research. *OT Pract*. 2002;7(7):9-10.
83. World Federation of Occupational Therapists. Guiding Principles for the Use of Evidence in Occupational…. WFOT. Published July 26, 2022. Accessed July 25, 2022. https://wfot.org/resources/guiding-principles-for-the-use-of-evidence-in-occupational-therapy
84. Hitch D, Nicola-Richmond K, Richards K, Stefaniak R. Student perspectives on factors that influence the implementation of evidence-based practice in occupational therapy. *JBI Evid Implement*. 2021;19(4):409-418. doi:10.1097/XEB.0000000000000285

85. Holm MB. Our mandate for the new millennium: Evidence-based practice. *Am J Occup Ther*. 2000;54(6):575-585.
86. Robertson L, Graham F, Anderson J. What Actually Informs Practice: Occupational Therapists' Views of Evidence. *Br J Occup Ther*. 2014;76:317-324. doi:10.4276/030802213X13729279114979
87. Kim SC, Stichler JF, Ecoff L, Brown CE, Gallo AM, Davidson JE. Predictors of Evidence-Based Practice Implementation, Job Satisfaction, and Group Cohesion Among Regional Fellowship Program Participants. *Worldviews Evid Based Nurs*. 2016;13(5):340-348. doi:10.1111/wvn.12171
88. Novak I, McIntyre S. The effect of education with workplace supports on practitioners' evidence-based practice knowledge and implementation behaviours. *Aust Occup Ther J*. 2010;57(6):386-393.
89. Bazyk S, Demirjian L, LaGuardia T, Thompson-Repas K, Conway C, Michaud P. Building capacity of occupational therapy practitioners to address the mental health needs of children and youth: A mixed-methods study of knowledge translation. *Am J Occup Ther*. 2015;69(6):6906180060p1-6906180060p10.
90. Fristedt S, Areskoug Josefsson K, Kammerlind AS. Factors Influencing the Use of Evidence-Based Practice among Physiotherapists and Occupational Therapists in Their Clinical Work. *Internet J Allied Health Sci Pract*. Published online 2016. doi:10.46743/1540-580X/2016.1553
91. Group NPA. Competency profile for physiotherapists in Canada. *Natl Physiother Advis Group*. Published online 2017. https://www.peac-aepc.ca/pdfs/Resources/Competency%20Profiles/Competency%20Profile%20for%20PTs%202017%20EN.pdf
92. Bernhardsson S, Johansson K, Nilsen P, Öberg B, Larsson MEH. Determinants of Guideline Use in Primary Care Physical Therapy: A Cross-Sectional Survey of Attitudes, Knowledge, and Behavior. *Phys Ther*. 2014;94(3):343-354. doi:10.2522/ptj.20130147
93. Upton D, Stephens D, Williams B, Scurlock-Evans L. Occupational Therapists' Attitudes, Knowledge, and Implementation of Evidence-Based Practice: A Systematic Review of Published Research. *Br J Occup Ther*. 2014;77:24. doi:10.4276/030802214X13887685335544

94. Saunders H, Gallagher-Ford L, Kvist T, Vehviläinen-Julkunen K. Practicing Healthcare Professionals' Evidence-Based Practice Competencies: An Overview of Systematic Reviews. *Worldviews Evid Based Nurs*. 2019;16(3):176-185. doi:https://doi.org/10.1111/wvn.12363
95. Thomas A, Zoubi FA, Mayo NE, et al. Individual and organizational factors associated with evidence-based practice among physical and occupational therapy recent graduates: A cross-sectional national study. *J Eval Clin Pract*. 2020;n/a(n/a). doi:https://doi.org/10.1111/jep.13518
96. Rochette A, Brousseau M, Vachon B, Engels C, Amari F, Thomas A. What occupational therapists' say about their competencies' enactment, maintenance and development in practice? A two-phase mixed methods study. *BMC Med Educ*. 2020;20. doi:10.1186/s12909-020-02087-4
97. Arumugam V, MacDermid JC, Walton D, Grewal R. Attitudes, knowledge and behaviors related to evidence-based practice in health professionals involved in pain management. *JBI Evid Implement*. 2018;16(2):107-118. doi:10.1097/XEB.0000000000000131
98. Hasani F, MacDermid JC, Tang A, Kho M, Alghadir AH, Anwer S. Knowledge, attitude and implementation of evidence-based practice among physiotherapists working in the kingdom of Saudi arabia: a cross-sectional survey. In: *Healthcare*. Vol 8. Multidisciplinary Digital Publishing Institute; 2020:354.
99. Graham F, Robertson L, Anderson J. N ew Z ealand occupational therapists' views on evidence-based practice: A replicated survey of attitudes, confidence and behaviours. *Aust Occup Ther J*. 2013;60(2):120-128.
100. Lyons C, Brown T, Tseng MH, Casey J, McDonald R. Evidence-based practice and research utilisation: Perceived research knowledge, attitudes, practices and barriers among Australian paediatric occupational therapists. *Aust Occup Ther J*. 2011;58(3):178-186.
101. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to Evidence-Based Practice Implementation in Physiotherapy: A Systematic Review and Meta-Analysis. *Int J Qual Health Care*. 2021;(mzab093). doi:10.1093/intqhc/mzab093

102. Pellerin MA, Lamontagne ME, Viau-Guay A, Poulin V. Systematic review of determinants influencing knowledge implementation in occupational therapy. *Aust Occup Ther J*. 2019;66(6):670-681. doi:10.1111/1440-1630.12612

103. Juckett LA, Wengerd LR, Faieta J, Griffin CE. Evidence-Based Practice Implementation in Stroke Rehabilitation: A Scoping Review of Barriers and Facilitators. *Am J Occup Ther Off Publ Am Occup Ther Assoc*. 2020;74(1):7401205050p1-7401205050p14. doi:10.5014/ajot.2020.035485

104. McCluskey A, Vratsistas-Curto A, Schurr K. Barriers and enablers to implementing multiple stroke guideline recommendations: a qualitative study. *BMC Health Serv Res*. 2013;13(1):323. doi:10.1186/1472-6963-13-323

105. Salbach NM, Guilcher SJ, Jaglal SB, Davis DA. Determinants of research use in clinical decision making among physical therapists providing services post-stroke: a cross-sectional study. *Implement Sci*. 2010;5(1):77. doi:10.1186/1748-5908-5-77

106. McEvoy MP, Williams MT, Olds TS. Development and psychometric testing of a trans-professional evidence-based practice profile questionnaire. *Med Teach*. 2010;32(9):e373-e380.

107. da Silva T, da Cunha Menezes Costa L, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. 2015;20(3):388-401. doi:10.1016/j.math.2014.10.009

108. Quartey J, Kwakye S. Barriers to evidence-based physiotherapy practice for stroke survivors in Ghana. *South Afr J Physiother*. 2018;74(1):1-11.

109. Ramírez-Vélez R, Correa-Bautista JE, Muñoz-Rodríguez DI, et al. Evidence-based practice: beliefs, attitudes, knowledge, and skills among Colombian physical therapists. *Colomb Médica*. 2015;46(1):33-40.

110. Ajzen I. From intentions to actions: A theory of planned behavior. In: *Action Control*. Springer; 1985:11-39.

111. Melas CD, Zampetakis LA, Dimopoulou A, Moustakis V. Evaluating the properties of the Evidence-Based Practice Attitude Scale (EBPAS) in health care. *Psychol Assess*. 2012;24(4):867.

112. Aarons GA. Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Ment Health Serv Res*. 2004;6(2):61-74.

113. Heiwe S, Kajermo KN, Tyni-Lenné R, et al. Evidence-based practice: attitudes, knowledge and behaviour among allied health care professionals. *Int J Qual Health Care*. 2011;23(2):198-209. doi:10.1093/intqhc/mzq083

114. Thomas A, Law M. Research Utilization and Evidence-Based Practice in Occupational Therapy: A Scoping Study. *Am J Occup Ther*. 2013;67(4):e55-e65. doi:10.5014/ajot.2013.006395

115. Iles R, Davidson M. Evidence based practice: a survey of physiotherapists' current practice. *Physiother Res Int*. 2006;11(2):93-103.

116. Lindström AC, Bernhardsson S. Evidence-based practice in primary care occupational therapy: a cross-sectional survey in Sweden. *Occup Ther Int*. 2018;2018.

117. Michie S, Johnston M, Abraham C, et al. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care*. 2005;14(1):26-33. doi:10.1136/qshc.2004.011155

118. Salbach NM, Jaglal SB %J J of evaluation in clinical practice. Creation and validation of the evidence-based practice confidence scale for health care professionals. 2011;17(4):794-800.

119. Boswell C, Ashcraft A, Long J, Cannon S, DiVito-Thomas P, Delaney T. Self-efficacy: changing the tide of evidence-based practice. *Worldviews Evid Based Nurs*. 2020;17(2):129-135.

120. Salbach NM, Jaglal SB, Williams JI. Reliability and validity of the Evidence-Based practice confidence (EPIC) scale. *J Contin Educ Health Prof*. 2013;33(1):33-40.

121. Lizarondo L, Grimmer-Somers K, Kumar S. A systematic review of the individual determinants of research evidence use in allied health. *J Multidiscip Healthc*. 2011;4:261-272. doi:10.2147/JMDH.S23144

122. Thomas Y, Dickson D, Broadbridge J, et al. Benefits and challenges of supervising occupational therapy fieldwork students: Supervisors' perspectives. *Aust Occup Ther J*. 2007;54:S2-S12.

123. Craik J, Rappolt S. Enhancing research utilization capacity through multifaceted professional development. *Am J Occup Ther*. 2006;60(2):155-164.

124. McQueen J, Nivison C, Husband V, Miller C. An investigation into the use of a journal club for evidence-based practice. *Int J Ther Rehabil*. 2006;13(7):311-317.

125. Stern P. Using journal clubs to promote skills for evidence-based practice. *Occup Ther Health Care*. 2008;22(4):36-53.

126. Bayley MT, Hurdowar A, Richards CL, et al. Barriers to implementation of stroke rehabilitation evidence: findings from a multi-site pilot project. *Disabil Rehabil*. 2012;34(19):1633-1638. doi:10.3109/09638288.2012.656790

127. Morris JH, Bernhardsson S, Bird ML, et al. Implementation in rehabilitation: a roadmap for practitioners and researchers. *Disabil Rehabil*. 2019;0(0):1-10. doi:10.1080/09638288.2019.1587013

128. Bennett S, Allen S, Caldwell E, et al. Organisational support for evidence-based practice: occupational therapists perceptions. *Aust Occup Ther J*. 2016;63(1):9-18. doi:10.1111/1440-1630.12260

129. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gómez JE. Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the literature: Evidence-based physiotherapy instruments. *J Eval Clin Pract*. 2014;20(6):767-778. doi:10.1111/jep.12180

130. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci*. 2014;9(1):118. doi:10.1186/s13012-014-0118-8

131. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

132. Shaneyfelt T, Baum KD, Bell D, et al. Instruments for Evaluating Education in Evidence-Based Practice: A Systematic Review. *JAMA*. 2006;296(9):1116-1127. doi:10.1001/jama.296.9.1116

133. Hrisos S, Eccles MP, Francis JJ, et al. Are there valid proxy measures of clinical behaviour? a systematic review. *Implement Sci*. 2009;4(1):37. doi:10.1186/1748-5908-4-37

134. Engelhard G. Introduction and Overview. In: *Invariant Measurement*. Routledge; 2012.

135. Romney W, Salbach N, Parrott JS, Deutsch JE. A knowledge translation intervention designed and implemented by a knowledge broker improved documented use of gait speed: a mixed-methods study. *J Geriatr Phys Ther*. 2020;43(3):E1-E10.

136. Lo L, Hebert D, Colquhoun H. Measuring practice gaps in the delivery of evidence-based seating assessments: a retrospective chart review. *Disabil Rehabil Assist Technol*. 2021;16(3):255-261. doi:10.1080/17483107.2019.1672814

137. Borgatti SP, Carboni I. On measuring individual knowledge in organizations. *Organ Res Methods*. 2007;10(3):449-462.

138. Spurlock D, Wonder AH. Validity and Reliability Evidence for a New Measure: The Evidence-Based Practice Knowledge Assessment in Nursing. *J Nurs Educ*. 2015;54(11):605-613. doi:10.3928/01484834-20151016-01

139. Tilson JK. Validation of the modified Fresno Test: assessing physical therapists' evidence based practice knowledge and skills. *BMC Med Educ*. 2010;10(1):38. doi:10.1186/1472-6920-10-38

140. Lewis LK, Williams MT, Olds TS. Development and psychometric testing of an instrument to evaluate cognitive skills of evidence based practice in student health professionals. *BMC Med Educ*. 2011;11(1):1-11.

141. Glegg SMN, Holsti L. Measures of knowledge and skills for evidence-based practice: A systematic review. *Can J Occup Ther Ott*. 2010;77(4):219-232.

142. Buchanan H, Siegfried N, Jelsma J. Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review: Evidence-based Practice Survey Instruments. *Occup Ther Int*. 2016;23(2):59-90. doi:10.1002/oti.1398

143. van de Mortel TF. Faking It: Social Desirability Response Bias in Self-report Research. *Aust J Adv Nurs*. Published online June 2008. Accessed March 18, 2021. https://search.informit.org/doi/abs/10.3316/INFORMIT.210155003844269

144. Stevenson K, Lewis M, Hay E. Do physiotherapists' attitudes towards evidence-based practice change as a result of an evidence-based educational programme? *J Eval Clin Pract*. 2004;10(2):207-217. doi:10.1111/j.1365-2753.2003.00479.x

145. Sick J. Assumptions and requirements of Rasch measurement. *Shiken JALT Test Eval SIG Newsl*. 2010;14(2):23-29.
146. Fernández-Domínguez JC, de Pedro-Gómez JE, Morales-Asencio JM, Bennasar-Veny M, Sastre-Fullana P, Sesé-Abad A. Health Sciences-Evidence Based Practice questionnaire (HS-EBP) for measuring transprofessional evidence-based practice: Creation, development and psychometric validation. Gagnier JJ, ed. *PLOS ONE*. 2017;12(5):e0177172. doi:10.1371/journal.pone.0177172
147. Van Laerhoven H, van der Zaag-Loonen HJ, Derkx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr*. 2004;93(6):830-835.
148. Davey HM, Barratt AL, Butow PN, Deeks JJ. A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *J Clin Epidemiol*. 2007;60(4):356-360.
149. Luo N, Li M, Chevalier J, Lloyd A, Herdman M. A comparison of the scaling properties of the English, Spanish, French, and Chinese EQ-5D descriptive systems. *Qual Life Res*. 2013;22(8):2237-2243. doi:10.1007/s11136-012-0342-0
150. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? *J Rehabil Med*. 2012;44(2):97-98. doi:10.2340/16501977-0938
151. Jamieson S. Likert scales: How to (ab) use them? *Med Educ*. 2004;38(12):1217-1218.
152. Bishop PA, Herron RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int J Exerc Sci*. 2015;8(3):297-302.
153. Andrich D. Rasch models for ordered response categories. *Encycl Stat Behav Sci*. Published online 2005.
154. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846-855. doi:10.1016/j.jval.2015.07.004
155. Young TA, Yang Y, Brazier JE, Tsuchiya A. The Use of Rasch Analysis in Reducing a Large Condition-Specific Instrument for Preference Valuation: The Case of Moving from AQLQ to AQL-5D. *Med Decis Making*. 2011;31(1):195-210. doi:10.1177/0272989X10364846

156. Andrich D. Rasch models. In: *International Encyclopedia of Education*. Pergamon; 2010:111-122.

157. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The First Stage of Developing Preference-Based Measures: Constructing a Health-State Classification Using Rasch Analysis. *Qual Life Res*. 2009;18(2):253-265.

158. Mayo N. *ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*. 1st edition. International Society for Quality of Life Research; 2015.

159. Feeny D. Preference-based measures: utility and quality-adjusted life years. *Assess Qual Life Clin Trials*. 2005;2:405-431.

160. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Med Care*. 2002;40(2):113-128.

161. Brazier J, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32). doi:10.3310/hta16320

162. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ*. 1999;8(1):41-51.

163. Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis*. 1987;40(6):593-600.

164. Guyatt GH, Rennie D. Users' guides to the medical literature. *Jama*. 1993;270(17):2096-2097.

165. Poissant L, Mayo NE, Wood-Dauphinee S, Clarke AE. The development and preliminary validation of a Preference-Based Stroke Index (PBSI). *Health Qual Life Outcomes*. Published online 2003:15.

166. Kuspinar A, Finch L, Pickard S, Mayo NE. Using existing data to identify candidate items for a health state classification system in multiple sclerosis. *Qual Life Res*. 2014;23(5):1445-1457. doi:10.1007/s11136-013-0604-5

167. Moga AM, Twells LK, Mayo NE. The prototype of a preference-based index of weight-related quality of life: demonstrating the possibilities. *Qual Life Res*. Published online May 24, 2022. doi:10.1007/s11136-022-03156-5

168. Green C, Brazier J, Deverill M. Valuing health-related quality of life. *Pharmacoeconomics*. 2000;17(2):151-165.

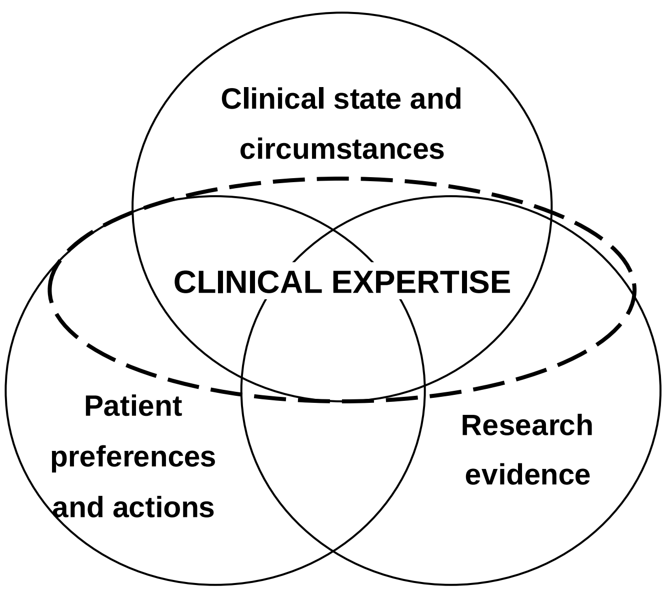
169. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *PharmacoEconomics*. 2017;35(1):21-31. doi:10.1007/s40273-017-0545-x
170. Weernink MGM, Janus SIM, van Til JA, Raisch DW, van Manen JG, IJzerman MJ. A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making. *Pharm Med*. 2014;28(4):175-185. doi:10.1007/s40290-014-0059-1
171. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403-413.
172. Green PE, Rao VR. Conjoint measurement-for quantifying judgmental data. *J Mark Res*. 1971;8(3):355-363.
173. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Evidence-Based Medicine E-Book: How to Practice and Teach EBM*. 5th edition. Elsevier Health Sciences; 2018.
174. Rycroft-Malone J, Seers K. What counts as evidence in evidence-based practice? *J Adv Nurs*. Published online 2004:10.
175. Upshur REG. Seven characteristics of medical evidence. *J Eval Clin Pract*. 2000;6(2):93-97. doi:10.1046/j.1365-2753.2000.00244.x
176. Higgs J, Titchen A. The nature, generation and verification of knowledge. *Physiotherapy*. 1995;81(9):521-530.
177. Maul A. On the Ontology of Psychological Attributes. *Theory Psychol*. 2013;23:752-769. doi:10.1177/0959354313506273
178. Goldenberg MJ. On evidence and evidence-based medicine: Lessons from the philosophy of science. *Soc Sci Med*. 2006;62(11):2621-2632. doi:10.1016/j.socscimed.2005.11.031
179. Bisel RS, Adame EA. Post-positivist/functionalist approaches. *Int Encycl Organ Commun*. Published online 2017:1-22.
180. Creswell JW. *A Concise Introduction to Mixed Methods Research*. SAGE publications; 2014.
181. O'donovan B, Price M, Rust C. Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teach High Educ*. 2004;9(3):325-335.
182. Chang H. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press; 2004.

Appendices

Appendix I: Figures of EBP models

Figure 1. Adaptation from Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72.

Figure 2. Adapted from Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *BMJ Evid-Based Med*. 2002;7(2):36-38



Appendix II: Subset of existing EBP measures used in rehabilitation

Title of the self-reported questionnaire	**EBP domains included (# of items)**	**Scale**
Philibert et al. (2003)	Attitudes towards research in practice (5 items)	6-point Likert-type scale ranging from Strongly Disagree (1) to Strongly Agree (6)
	Sources of knowledge to guide practice (7 items)	6-point Likert-type scale ranging from Always (6) to Never (1)
Evidence-Based Practice Questionnaire (EBPQ) also called the *revised Upton* (Upton and Upton, 2006)	Knowledge and skills associated with EBP (14 items)	7-point Likert-type scale ranging from 1 (poor) to 7 (best)
	Attitudes towards EBP (4 items)	7-point Likert-type scale, ranging from strongly disagree to strongly agree
	Practice of EBP (6 items)	7-point Likert-type scale, ranging from never to frequently
Barriers to Research Utilization Scale (BARRIERS) (Funk et al., 1991)	Perceived barriers to research utilization (28 items): (i) the adopter (values, skills and awareness); (ii) the organisation (setting); (iii) the innovation (qualities of the research); and (iv) the communication (presentation and accessibility of the research).	4-point Likert-type scale, ranging from 1 ('to no extent') to 4 ('to a great extent')
Barriers and Attitudes to Research in the Therapies (BART) (Metcalfe et	Perceived importance of research (7 items)	Binary, agree (-1) or disagree (1), total score ranging from -7 to 7
	Perceived barriers to research (22 items)	Binary, agree (-1) or disagree (1),

al., 2001)		total score ranging from -22 to 22
Edmonton Research Orientation Survey (EROS) (Pain et al., 1996)	Research utilization (38 items): (i) valuing research; (ii) research involvement; (iii) being at the leading edge; and (iv) evidence-based practice	5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree)
Knowledge, Attitudes, and Behavior questionnaire (KAB) (Stronge and Cahill, 2012)	Self-rated knowledge (5 items)	6-point scale from 1 (strongly disagree) to 6 (strongly agree)
	Attitudes to EBP (6 items)	6-point scale from 1 (strongly disagree) to 6 (strongly agree)
	Current use of EBP (6 items)	5-point scale from 1 (never) to (5) every day
	future use of EBP (9 items)	6-point scale from 1 (not at all) to 6 (completely)
The Knowledge, Attitudes, and Practices survey (KAP) (Van Mullem et al., 1999)	Knowledge of activities related to utilizing research (5 items)	3-point Likert-type scale, including low (1), moderate (2) and high (3)
	Willingness to engage in activities related to utilizing research (5 items)	3-point Likert-type scale, including low (1), moderate (2) and high (3)
	Ability to perform activities related to utilizing research (5 items)	3-point Likert-type scale, including low (1), moderate (2) and high (3)
EBP and clinical effectiveness questionnaire (Upton and Lewis, 1998)	Perceived knowledge of EBP and clinical effectiveness (1 item)	5-point visual analogue scale ranging from 1 ('I know a Great deal about clinical Effectiveness and EBP') to 5 (' I know very little about clinical effectiveness and EBP')
	Perceived knowledge of individual component skills of EBP (5 items)	7-point Likert-type scale, ranging from 1 (poor) to 7 (best).

	Attitudes to EBP (# of items not reported)	Sliding scale between 'EBP is a waste of time' to 'EBP is fundamental to professional practice' Incomplete information. Ranges from agreed to disagreed with statements.
	Frequency of completing EBP steps (# of items not reported)	Incomplete information. Ranges from never to frequently.
	Barriers and solutions to implementing EBP (# of items not reported)	Incomplete information. Rate level of agreement with perception of barriers and solutions to EBP.
Health Sciences Evidence Based Questionnaire (HS-EBP) (Fernandez-Domínguez et al., 2017)	Beliefs and attitudes (12 items)	10-point Likert-type scale ranging from 1 (lowest level of agreement) to 10 (highest level of agreement)
	Results from scientific research (14 items)	10-point Likert-type scale ranging from 1 (lowest degree of frequency) to 10 (highest degree of frequency)
	Development of professional practice (10 items)	10-point Likert-type scale ranging from 1 (lowest degree of frequency) to 10 (highest degree of frequency)
	Assessment of results (12 items)	10-point Likert-type scale ranging from 1 (lowest degree of frequency) to 10 (highest degree of frequency)
	Barriers/Facilitators (12 items)	10-point Likert-type scale ranging from 1 (lowest level of agreement) to 10 (highest level of agreement)

Al Zoubi et al. (2018)	Self-use of EBP (9 items)	5-point Likert-type scale ranging from “never” to “more than 10 times a month”
	EBP activities (7 items)	5-point Likert-type scale ranging from “never” to “daily”
	Attitudes towards EBP (17 items)	5-point Likert-type scale ranging from “strongly disagree” to “strongly agree”
	EBP self-efficacy (9 items)	Scale from 0 to 10 representing 0–100%
	Knowledge of EBP (11 items)	5-point Likert-type scale ranging from “never heard the term” to “understand and could explain to others”
	EBP resources (17 items)	5-point Likert-type scale ranging from “strongly disagree” to “strongly agree”
EBP, evidence-based practice; BART, Barriers and Attitudes to Research in the Therapies; EBPQ, Evidence-Based Practice Questionnaire; EROS, Edmonton Research Orientation Scale; KAB, Knowledge, Attitudes, Behaviour; KAP, Knowledge Attitudes and Practice of Research		

Chapter 2: Rationale and objectives

Rationale

Evidence-based practice (EBP) is not only a core professional competency for occupational therapist and (OTs) and physical therapists (PTs) but an expectation from consumers and funders who expect and deserve quality services based on best available evidence[1]. EBP is defined as an approach to clinical decision-making (CDM) that combines the best available research evidence, clinical expertise, and patients' preferences, within the context of available resources[2–4]. Robust measurement practices are essential to determine if clinicians are engaging in the EBP process, and to identify the factors related to EBP that should be improved or maintained. The measurement of EBP can serve to identify research and practice needs, better support clinicians and ultimately, improve quality of care. Despite the growing emphasis on robustly measuring EBP, there are many challenges including: (1) the shortage of comprehensive questionnaires representing the most salient individual *and* organizational EBP domains; (2) the inappropriate treatment of ordinal data as interval-level data; (3) the fragmented interpretation of EBP due to "profile-type" measures which summarize results of domains independently; and (4) the unknown relative importance of the domains that influence EBP and how these can be quantitatively captured in a measure.

Research Objectives

The overall objective of this PhD thesis was to take the necessary steps towards developing a brief, multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into clinical decision-making to be used in studies of evidence-based practice.

To operationalize the overarching aim of my doctoral work, I conducted four research projects, described in five manuscripts that compose this doctoral dissertation, each with the following objectives:

Study 1

The global aim of this study was to describe the methods, results, recommendations, reported challenges and areas for future practice across systematic reviews on EBP measures in healthcare.

> *Manuscript 1: Quality, Methods, and Recommendations of Systematic Reviews on Measures of Evidence-Based Practice: An Umbrella Review*

> *Manuscript 2: Challenges and Future Directions in the Measurement of Evidence-based Practice: Qualitative Analysis of Umbrella Review Findings*

Study 2

The global aim of this study was to describe the prototype development of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific study objectives were to: (1) identify candidate items best reflecting the most salient EBP domains for OTs and PTs across Canada from existing measures; (2) estimate the extent to which the prototype PIRE-CDMI "behaves" coherently across characteristics of the sample; and (3) estimate the extent to which the prototype PIRE-CDMI provides information that is comparable to that from other EBP measures.

> *Manuscript 3: Identifying Candidate Items for a Prototype Index on Propensity to Integrate Research Evidence into Clinical Decision-making in Rehabilitation*

Study 3

The global aim of this study was to contribute evidence for the clarity and interpretability of items and response options for a new bilingual (English and French) measure, the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). Specifically, the primary objective of this research was to qualitatively review and revise the included items in the prototype index in English and French The secondary objective was to estimate the equivalency of response option labels in both languages.

> <u>*Manuscript 4*</u>*: Improving the Clarity and Interpretability of Items in a Bilingual Index of Propensity to Integrate Research Evidence into Clinical Decision-making in Rehabilitation*

Study 4

The global aim of this study was to develop a scoring algorithm for the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The primary objective was to estimate the part-worth utilities and relative importance of five key dimensions to the propensity to integrate research evidence into clinical decision-making, information that will be used in the scoring algorithm. The secondary objective was to contribute evidence of the ability of the PIRE-CDMI to distinguish between known groups of OTs and PTs and to compare with a global self-rating of EBP.

> <u>*Manuscript 5*</u>*: Propensity to Integrate Research Evidence into the Decision-making Process in Rehabilitation: Development of a Weighted Algorithm Using a Best-Worst Scaling Choice Exercise*

References

1. CAOT, ACOTRO, ACOTUP, PAC. Joint position statement on evidence-based occupational therapy. *Can J Occup Ther*. 2009;66(5):267-269. https://www.caot.ca/document/3697/J%20-%20Joint%20Position%20Statement%20on%20Evidence%20based%20OT.pdf

2. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1

3. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Evidence-Based Medicine E-Book: How to Practice and Teach EBM*. 5th edition. Elsevier Health Sciences; 2018.

4. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014

Chapter 3.1: Manuscript 1

Quality, Methods, and Recommendations of Systematic Reviews on Measures of Evidence-based Practice: An Umbrella Review

Jacqueline Roberge-Dao[1,2], Lauren A. Maggio[3], Marco Zaccagnini[1,2], Annie Rochette[2,4], Keiko Shikako[1,2], Jill Boruff[5], Aliki Thomas[1,2,6]

[1] School of Physical and Occupational Therapy, McGill University, Montréal, Canada
[2] Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Canada
[3] Medicine and Health Professions Education, Uniformed Services University, Bethesda, USA
[4] L'École de réadaptation de l'Université de Montréal, Montréal, Canada
[5]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Canada
[6]The Institute of Health Sciences Education, McGill University, Montréal, Canada

Corresponding author: Aliki Thomas, aliki.thomas@mcgill.ca, 3654 Promenade Sir William Osler, Montreal, QC, Canada, H3G-1Y5, 514-398-4496

Published in *JBI Evidence Synthesis* (2022)
doi: 10.11124/JBIES-21-00118

Quality, methods, and recommendations of systematic reviews on measures of evidence-based practice: an umbrella review

Jacqueline Roberge-Dao[1,2] • Lauren A. Maggio[3] • Marco Zaccagnini[1,2] • Annie Rochette[2,4] • Keiko Shikako-Thomas[1,2] • Jill Boruff[5] • Aliki Thomas[1,2]

[1]School of Physical and Occupational Therapy, McGill University, Montréal, QC, Canada, [2]Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, QC, Canada, [3]Medicine and Health Professions Education, Uniformed Services University of the Health Sciences, Bethesda, MD, USA, [4]School of Rehabilitation, Université: de Montréal, Montréal, QC, Canada, and [5]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, QC, Canada

ABSTRACT

Objectives: The objective of the review was to estimate the quality of systematic reviews on evidence-based practice measures across health care professions and identify differences between systematic reviews regarding approaches used to assess the adequacy of evidence-based practice measures and recommended measures.

Introduction: Systematic reviews on the psychometric properties of evidence-based practice measures guide researchers, clinical managers, and educators in selecting an appropriate measure for use. The lack of psychometric standards specific to evidence-based practice measures, in addition to recent findings suggesting the low methodological quality of psychometric systematic reviews, calls into question the quality and methods of systematic reviews examining evidence-based practice measures.

Inclusion criteria: We included systematic reviews that identified measures that assessed evidence-based practice as a whole or of constituent parts (eg, knowledge, attitudes, skills, behaviors), and described the psychometric evidence for any health care professional group irrespective of assessment context (education or clinical practice).

Methods: We searched five databases (MEDLINE, Embase, CINAHL, PsycINFO, and ERIC) on January 18, 2021. Two independent reviewers conducted screening, data extraction, and quality appraisal following the JBI approach. A narrative synthesis was performed.

Results: Ten systematic reviews, published between 2006 and 2020, were included and focused on the following groups: all health care professionals (n = 3), nurses (n = 2), occupational therapists (n = 2), physical therapists (n = 1), medical students (n = 1), and family medicine residents (n = 1). The overall quality of the systematic reviews was low: none of the reviews assessed the quality of primary studies or adhered to methodological guidelines, and only one registered a protocol. Reporting of psychometric evidence and measurement characteristics differed. While all the systematic reviews discussed internal consistency, feasibility was only addressed by three. Many approaches were used to assess the adequacy of measures, and five systematic reviews referenced tools. Criteria for the adequacy of individual properties and measures varied, but mainly followed standards for patient-reported outcome measures or the Standards of Educational and Psychological Testing. There were 204 unique measures identified across 10 reviews. One review explicitly recommended measures for occupational therapists, three reviews identified adequate measures for all health care professionals, and one review identified measures for medical students. The 27 measures deemed adequate by these five systematic reviews are described.

Conclusions: Our results suggest a need to improve the overall methodological quality and reporting of systematic reviews on evidence-based practice measures to increase the trustworthiness of recommendations and allow comprehensive interpretation by end users. Risk of bias is common to all the included systematic reviews, as the quality of primary studies was not assessed. The diversity of tools and approaches used to evaluate the

Correspondence: Aliki Thomas, aliki.thomas@mcgill.ca

The authors declare no conflict of interest.

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the U.S. Government.

DOI: 10.11124/JBIES-21-00118

adequacy of evidence-based practice measures reflects tensions regarding the conceptualization of validity, suggesting a need to reflect on the most appropriate application of validity theory to evidence-based practice measures.

Systematic review registration number: PROSPERO CRD42020160874

Keywords: evidence-based practice; instruments; psychometrics; systematic review; umbrella review

JBI Evid Synth 2022; 20(4):1004–1073.

Introduction

Calls for quality of care, improved patient outcomes, and efficient use of resources have prompted the rapid growth of the evidence-based practice (EBP) movement.[1-3] Evidence-based practice is broadly defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."[4(p.71)] The teaching of EBP largely focuses on the knowledge and skills associated with five steps: ask, acquire, appraise, apply, and assess.[5] This well-established operationalization of EBP, popularized through the Sicily statement, is known as the five-step cycle or the process of EBP.[5-7]

Following an EBP educational activity or clinical intervention, it is best practice for educators, clinical managers, and researchers to ascertain the effectiveness of the activity. The past two decades have seen rapid developments in EBP assessments (eg, tests, questionnaires, chart audits) with variable purposes, scale constructions, and practicality.[8] Some assessments focus on one step of the EBP process (eg, critical appraisal), while others cover many steps.[5] Similarly, measures may assess a set of EBP domains (eg, knowledge, skills, attitudes, behaviors), while others may only assess a single domain.[8] The wide selection of EBP assessment approaches and many different measurement characteristics and configurations have made selecting a robust and relevant measure time-consuming and challenging. The choice of an EBP assessment must be carefully thought out and applicable to the assessment context to avoid producing misleading conclusions.[9,10]

Systematic reviews of EBP measures can guide EBP researchers and educators in selecting an appropriate and robust measure. High-quality systematic reviews on EBP measures should provide comprehensive information on psychometric evidence and clear and trustworthy recommendations of the most suitable measures to use in a particular assessment context.[11] However, the trustworthiness of the recommendations is contingent on the methodological quality of the review, and recent studies have appraised systematic reviews of measurement properties as being of poor quality.[12,13] Additionally, the recommendations of systematic reviews largely depend on the measurement standards used to assess the adequacy of measures. Thus, different methods of assessing the quality of measures may result in diverging review findings.[14] Various measurement standards, such as the Terwee criteria and the Standards for Educational and Psychological Testing, have been developed to assess the quality of patient-reported or clinician-reported outcome measures.[12-14] As EBP is an approach to clinical decision-making, the proximal outcomes assessed by EBP measures are clinician-related and constitute a measurement target population that is not exactly patient-reported or clinician-reported outcomes. It has not yet been established which measurement standards are most appropriate for measures of EBP. The lack of consensus-based psychometric standards specific to EBP measures, in addition to findings suggesting the low methodological quality of psychometric systematic reviews, call into question the quality and methods of systematic reviews examining EBP measures.

This umbrella review seeks to compare psychometric systematic reviews of EBP measures across health care professions.

Review objectives

The specific aims of this review are to i) estimate the quality of psychometic systematic reviews of EBP measures; and ii) identify differences between these reviews regarding a) approaches used to assess the adequacy of EBP measures and b) recommendations for EBP measures.

Inclusion criteria

Participants

This umbrella review considered reviews reporting on health care professionals, including, but not limited to, physicians, nurses, occupational therapists, physiotherapists, and speech-language pathologists, as well as health care professional students. We did not restrict inclusion based on age, sex, or level of expertise.

Phenomena of interest

The phenomena of interest for this review were measures assessing EBP as a whole or of EBP constituent parts, including, but not limited to, EBP domains of knowledge, attitudes, self-efficacy, skills, behaviors, or the EBP steps: ask, acquire, appraise, apply, and assess. For this study, a "measure" was defined as a method to collect information, and includes questionnaires, instruments, and other measurement methods. We accepted similar terms to EBP (eg, evidence-informed) if the measurement focus was the extent to which clinicians integrate research evidence into their clinical reasoning. We excluded reviews that i) did not report on specific EBP measures and psychometric properties; ii) examined measures of "research utilization," as this is conceptually distinct from EBP (although often used interchangeably), and the focus of our review is on the broader aggregate process of EBP; and iii) examined measures assessing more peripheral concepts to EBP (eg, critical thinking, clinical reasoning, continuing education).

Context

This umbrella review considered reviews examining EBP measures in educational (ie, university) or clinical settings (eg, clinical practice, clinical environment). We excluded reviews that examined measures for a specific area of clinical practice (eg, stroke) as the measures would be too diagnosis-specific for our analysis of EBP as a broad construct.

Types of studies

We included reviews that were labeled as systematic reviews and that identified and reported the psychometric properties of EBP measures. We excluded protocols, theses, and conference papers.

Methods

Umbrella reviews compare and contrast published systematic reviews, with the aim to provide an overall examination of a body of information available for a given topic.[15,16] This review was conducted following the JBI methodology for umbrella reviews and according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[16,17] No institutional review board approval was necessary. A preliminary search of several databases (JBI Database of Systematic Reviews and Implementation Reports, the Cochrane Database of Systematic Reviews, PROSPERO, CINAHL, and MEDLINE) was conducted and revealed that no umbrella reviews on the topic of EBP measures in health care had been published or were in progress. The a priori protocol for this review was registered on PROSPERO (CRD42020160874).

Modifications from the protocol

The initial research question in the protocol was modified during data extraction when the research team found that only one systematic review explicitly recommended EBP instruments. As per the initial question, comparing the characteristics and measurement properties of EBP instruments from that single systematic review would not have contributed to the literature. Thus, we reframed our research question with the systematic review as the unit of analysis. We focused on the quality of reviews and approaches to assessing adequacy of measures as this information influences the results and recommendations regarding EBP measures reported in the systematic reviews. Consequently, we refined our inclusion criteria to solely include articles that were labeled as systematic reviews to focus on this form of evidence synthesis, which has high standards for methodological rigor. Further, the outcomes stated in the protocol no longer applied given the change in research question.

In this manuscript, we chose not to address the second research question in our protocol as it will be addressed in a future manuscript that will report on a secondary qualitative analysis of the main trends, gaps, and reported challenges regarding measurement of EBP across systematic reviews.

Systematic reviews on measures of research utilization were excluded from the search because research utilization is a subset of EBP and does not, in fact, include all five EBP steps as was initially stated in the protocol. Although this does not represent a deviation in methods from the protocol because there was never a restriction on level of expertise (eg, students, practicing clinicians), our protocol did not explicitly distinguish education and clinical settings under "context" as we have reported in this review.

Search strategy

We used a three-step search strategy approach. First, we conducted a preliminary search of the selected databases to identify relevant keywords and index terms. Second, we searched MEDLINE (Ovid), Embase (Ovid), CINAHL (EBSCO), PsycINFO (Ovid), and ERIC (EBSCO) on January 18, 2021, to identify systematic reviews aligning with the inclusion criteria. We developed specific search strategies for each database in collaboration with a health sciences librarian (JB). Our search strategy combined subject headings and keywords related to EBP, measures, and health care professionals with no restrictions on publication date or language. Complete search strategies for all databases are available in Appendix I. Results were limited to systematic reviews using the Canadian Agency for Drugs and Technologies in Health systematic review filter.[18] Reviews published in English or French (in alignment with the research team's language proficiencies) were eligible for inclusion. Third, we manually searched the reference lists of all included studies to identify any studies that may have been missed.

We did not consider gray literature in our search as we were only focused on retrieving systematic reviews, which are unlikely to be found in the gray literature.

Study screening and selection

We imported the citations into EndNote v.X9.3.3 (Clarivate Analytics, PA, USA), removed duplicates, and uploaded the remaining citations into Covidence Systematic Review Software (Veritas Health Innovation, Melbourne, Australia).

We conducted a pilot selection process on a randomly selected 5% of articles and made minor revisions to the selection criteria to optimize agreement. Two reviewers (JRD and MZ) independently screened titles and abstracts for eligibility. Then, they independently reviewed the full text of potentially relevant studies. A third reviewer (AT) arbitrated disagreements that could not be resolved through consensus.

Assessment of methodological quality of systematic reviews

Two reviewers independently appraised included systematic reviews using the JBI critical appraisal checklist for systematic reviews and research synthesis.[19] Discrepancies were resolved through discussion. For each criterion, studies were rated as met (1), unclear (0.5), not met (0), or not applicable (N/A). Item 9, which pertains to the likelihood of publication bias, was omitted from the critical appraisal criteria for this review as included studies did not examine effectiveness.[20,21] We decided a priori not to exclude any reviews based on quality assessment but to present all quality appraisal results. This is a deviation from the JBI methodology, which recommends setting a cut-off score for quality.[22] As this umbrella review focuses, in part, on examining the quality of systematic reviews, we deemed it important to present data on all included reviews and allow readers to judge whether the information from specific systematic reviews is valuable.

Data collection

The data extraction form was developed and customized for the purposes of this research and was organized in a Microsoft Excel spreadsheet (Redmond, Washington, USA). Two reviewers (JRD, MZ) independently piloted the form on a randomly selected 10% of included studies for minor revisions to enhance clarity, relevance, and completeness. The pilot process included applying the critical appraisal tool. Disagreements between the reviewers were resolved through discussion and consensus.

Two reviewers independently (JRD, MZ) extracted data from eligible reviews for information on the i) characteristics of reviews (including methods); ii) amount of detail provided on measure characteristics, psychometric properties, and adequacy assessment results; iii) method of assessing the adequacy of measures; and iv) recommended EBP measures, including the characteristics of these measures. The data extraction form is provided in Appendix II.

The full list of included measures was missing for three reviews; we contacted these authors to obtain this information. We mapped the depth and breadth of reporting onto 11 properties reported in the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews of outcome measurement instruments.[23] For the purposes of this review, a recommended measure is a measure that was either explicitly recommended for use or was reported by the authors as being of high quality, adequate, and/or suitable for use for the specified population.[24]

Data summary

We used descriptive statistics to synthesize findings from each systematic review, such as the characteristics of the systematic reviews and the recommended measures. We compiled the names of all the measures identified by reviews alongside the referenced primary studies to control for duplicates in tabular format.[25] We conducted a narrative synthesis because statistical pooling of psychometric properties was not the aim of this review. Narrative synthesis involved the juxtaposition of findings from individual systematic reviews (eg, the type of psychometric property assessed, the breadth of reporting of psychometric properties) using a textual approach (eg, keywords) to describe and summarize the findings.[26,27]

Results

Study inclusion

The results of the search and selection of articles are included in Figure 1.[17] After removing duplicates, our search identified 3572 articles. Eighty-seven full-text articles met the inclusion criteria and were screened for eligibility, and 10 (11.5%) were retained for the full umbrella review (including one article found from reference checking the nine included articles). The excluded articles (n = 78) and reasons for exclusion are provided in Appendix III. There was substantial[28,29] inter-rater agreement for the inclusion process at the title and abstract level, with 99% proportionate agreement (K = 0.67, CI 95% [0.57 to 0.78]) at the screening phase (n = 3393 articles) after the calibration exercise[30]; more information on inter-rater agreement can be found in Appendix IV.

Characteristics of included studies

Table 1 includes descriptive information on the included systematic reviews. Five reviews aimed to identify, describe, and appraise EBP measures,[31-35] four reviews aimed to identify and describe EBP measures,[36-39] and one review focused on identifying the content of EBP interventions and how the effect was measured.[40] All reviews examined a range of psychometric properties instead of focusing on a single property (eg, content validity). The population scope differed across reviews from mixed to specific health care professional groups with varying specificity. The inclusion criteria for EBP measures regarding the type of assessment and EBP domains covered also varied.

The systematic reviews included on average 42 EBP measures per review (range 8 to 160). We identified 204 unique EBP measures across all 10 reviews. The most frequently identified measure was the Evidence Based Practice Questionnaire, which was included in half of the systematic reviews,[31-33,36,37] followed by the Fresno test[34,35,39,40] and the Taylor *et al.* questionnaire,[34,38-40] which were each reported by four systematic reviews. The list of all identified measures by systematic review is available in Appendix V.

Methodological quality of systematic reviews

Critical appraisal

Table 2 provides a breakdown of the critical appraisal scores. The rationale for ratings is available in Appendix VI.

Appendix VII provides additional information regarding methodological aspects of reviews. None of the reviews assessed the methodological quality of included primary studies or adhered to a methodological guideline for systematic reviews. One registered a protocol prospectively,[36] two included a PRISMA checklist,[39,40] four reported adhering to PRISMA reporting guidelines,[33,36,37,39] and nine included a search outcome flowchart.[31,33-40] Regarding the search strategy, three reviews built and conducted the search with a librarian or information specialist,[36,39,40] five reported the detailed syntax for at least one database,[33,36-39] six reported having two independent reviewers for study eligibility at both title/ abstract and full-text level,[31,34-38] six performed reference checking of included articles,[31,32,34,38-40] and nine provided reasons for article exclusion.[31,33-40]

Reporting of measurement properties across systematic reviews

Most reviews reported psychometric information for all included measures, except two reviews,[34,40] which reported psychometric information for a subset of higher-quality measures. The inclusion and labeling of psychometric properties varied across reviews. Internal consistency was the most frequently discussed psychometric property with all but two reviews[34,40] reporting Cronbach's alpha scores. Nine reviews[31-35,37-40] reported on separate kinds of validity (eg, construct, content, criterion), while one[36] reported sources of validity evidence (eg, content, response process, internal structure). Three reviews addressed feasibility of included measures (eg, administration and scoring time, training requirements, respondent

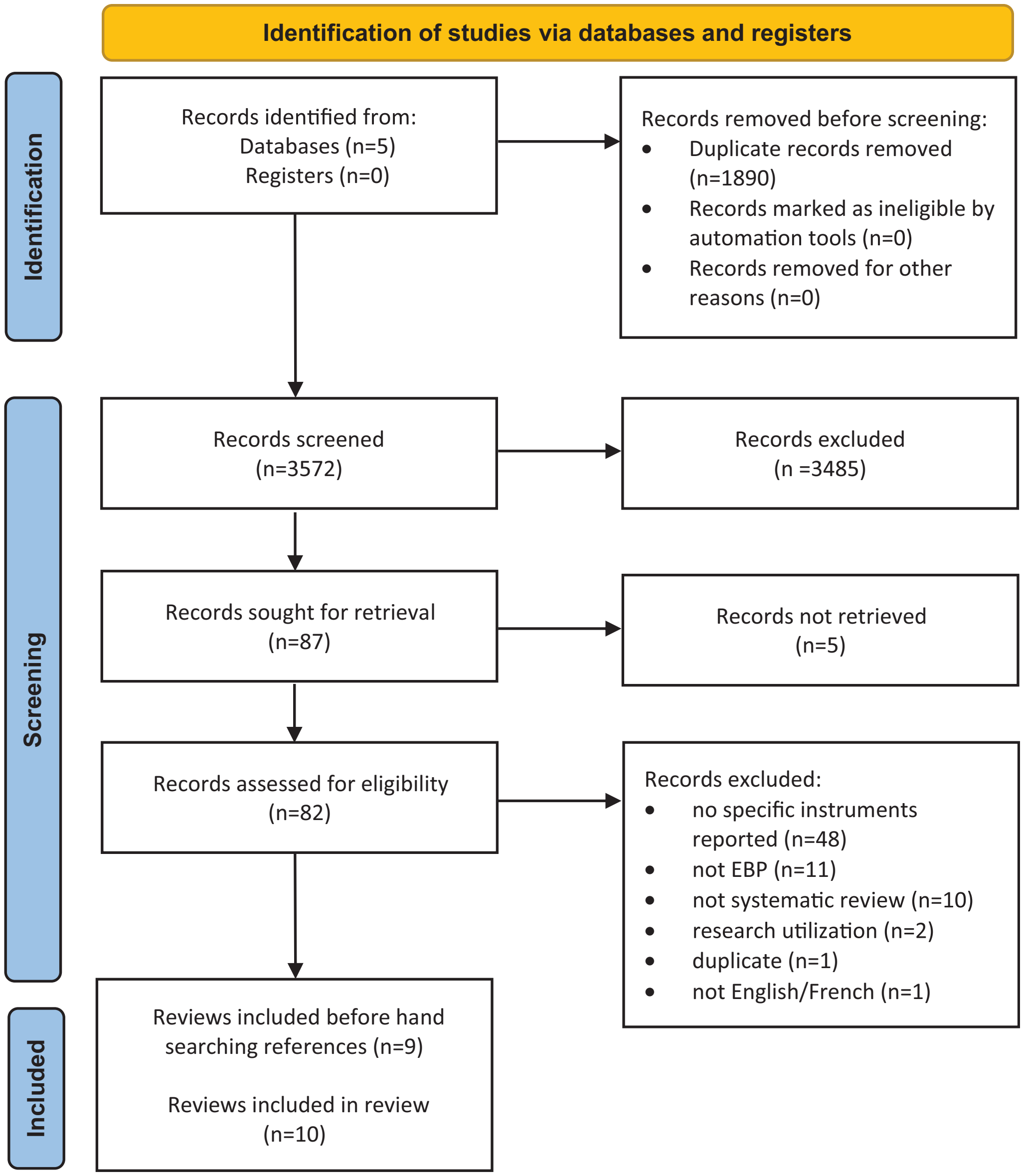


Figure 1: Search results and review selection and inclusion process[17]

burden)[31,34,36]; other reviews may have sporadically reported the mode of administration and completion time of measures. Appendix VIII provides additional information on the depth and breadth of reporting across reviews.

For each psychometric property, the amount of detail provided on the primary study design, participants, sample size, analyses, and results varied across and within reviews. For example, in one review,[39] inter-rater reliability information ranged from a standalone score to describing the raters, analyses, and scores. The reporting of information related to content validity varied the most across reviews and ranged from reporting whether content validity was "clearly, partially or not stated" in primary studies; to describing it as "established" to providing a keyword to describe the methods (eg, expert panel); to reporting the item review methods, participants, analysis, and results. Six reviews provided a description of the intended

Table 1: Description of included systematic reviews (n = 10)

Systematic review	Country	Population	EBP context	Inclusion criteria for EBP measures		Number of included studies	Number of included measures	Number of identifiable measures
				Type of assessment	EBP domains			
Albarqouni *et al.*[40] (2018)	Australia	All HCPs	CP + Education	Instruments reported in the included studies (controlled studies examining effect of EBP educational interventions)	EBP	85	24[a]	6
Belita *et al.*[36] (2020)	Canada	Nurses	CP	Measures	EIDM K, S, A, B	103	35	35
Buchanan *et al.*[31] (2016)	South Africa	OTs	CP + Education	Survey instruments	EBP K, S, A, B	35	34	34
Fernández-Domínguez *et al.*[38] (2014)	Spain	PTs	CP	Measures	EBP	24	24	24
Glegg and Holsti[32] (2010)	Canada	OTs	CP	Self-report and competence-based measures	EBP K, S	NR	15	15
Kumaravel *et al.*[39] (2020)	UK	Medical students	Education	Objective (non-self-reported) tools	EBM	12	12	12
Leung *et al.*[33] (2014)	Australia	Nurses	CP	Tools	EBP K, S, A,	59	16	16
Oude Rengerink *et al.*[37] (2013)	The Netherlands	All HCP	CP	All existing methods	EBP behavior	172	160[a]	82
Shaneyfelt *et al.*[34] (2006)	USA	All HCP	CP + Education	Instrument or strategy	EBP K, S, A, B	115	104[a]	20
Thomas and Kreptul[35] (2015)	Canada	Fam med residents	Education	Competence tests	EBM	11	8	8
Range						[11-172]	[8-160]	[6-82]
Mean						68.4	43.2	25.2
Median						59	24	18

A, attitudes; B, behaviors; CP, clinical practice; EBM, evidence-based medicine; EBP, evidence-based practice; EIDM, evidence-informed decision-making; fam med, family medicine; HCP, health care professional; K, knowledge; OT, occupational therapist; PT, physical therapist; S, skills; NR, not reported
[a]Did not report identifiable information on all included measures (ie, name of measure and/or source study)

Table 2: Critical appraisal of included systematic reviews

Citation (year)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Score
Albarqouni *et al.*[40] (2018)	1	1	1	0	0	0	1	1	N/A	1	1	7
Belita *et al.*[36](2020)	1	1	1	1	0	0	1	1	N/A	0	1	7
Buchanan *et al.*[31] (2016)	0	1	1	1	0	0	0	1	N/A	0.5	1	5.5
Fernández-Domínguez *et al.*[38] (2014)	1	1	1	1	0	0	1	0	N/A	0	1	6
Glegg and Holsti[32] (2010)	1	0	0	0	0	0	0	1	N/A	0	1	3
Kumaravel *et al.*[39] (2020)	1	1	0	0.5	0	0	1	1	N/A	1	1	6.5
Leung *et al.*[33] (2014)	1	1	1	0.5	0	0	1	1	N/A	1	1	7.5
Oude Rengerink *et al.*[37] (2013)	0	1	1	1	0	0	1	1	N/A	0	0	5
Shaneyfelt *et al.*[34] (2006)	0	1	1	1	0	0	1	1	N/A	1	1	7
Thomas and Kreptul[35] (2015)	0	0	0	1	0	0	1	0	N/A	0	0	2
Total %	60	80	70	70	0	0	80	80	N/A	45	80	

1, yes; 0, no; 0.5, unclear; N/A, not applicable; JBI critical appraisal checklist for systematic reviews and research syntheses.
Q1 Is the review question clearly and explicitly stated?
Q2 Were the inclusion criteria appropriate for the review question?
Q3 Was the search strategy appropriate?
Q4 Were the sources and resources used to search for studies adequate?
Q5 Were the criteria for appraising studies appropriate?
Q6 Was critical appraisal conducted by two or more reviewers independently?
Q7 Were there methods to minimize errors in data extraction?
Q8 Were the methods used to combine studies appropriate?
Q9 Was the likelihood of publication bias assessed?
Q10 Were recommendations for policy and/or practice supported by the reported data?
Q11 Were the specific directives for new research appropriate?

purpose of measures.[32,34,36,38-40] The only review to mention the theoretical basis of measures was Fernández-Domínguez *et al.*,[38] who described it as "clearly, partially or not stated"[(p.775)] in studies. No review discussed the measurement model (ie, formative or reflective constructs). The Belita *et al.* review[36] was the only review to report full raw data (ie, study methods, analyses, and results) from the primary studies for all included properties. One review did not report actual scores for psychometric properties.[40] All other reviews presented scores for some, but not all, included properties.

Four reviews classified psychometric evidence by measure[32,34,39,40] and six by primary study.[31,33,35-38] For the studies that classified by measure, there was no distinction of the sample population associated with the psychometric information as the evidence was compiled from various studies. For measures that were adaptations of other measures, three reviews[31,32,37] reported the original measure's psychometric information. The included systematic reviews were descriptive in nature and did not conduct pooled estimates of measurement properties. Only Belita *et al.*[36] reported pooled estimates of Cronbach's alpha for some measures; however, the pooling methods were not described.

Adequacy assessment of evidence-based practice measures across systematic reviews

Among the eight reviews that described the assessment of measures, approaches varied considerably as shown in Table 3.[41-47] Six reviews[31-34,39,40] determined the adequacy of a measure based on various criteria for determining the strength of psychometric evidence (see Appendix IX for additional information on the criteria from three reviews) while two[37,38] judged the quality of a measure based on the availability of psychometric testing in the literature. Seven reviews synthesized the findings of individual psychometric properties into an overall rating of quality for measures.[31-34,37,39,40] Five reviews presented assessment results at both the psychometric property and measure levels.[31,32,34,37,39] Three reviews used the Shaneyfelt *et al.* approach,[34,39,40] and five reviews referenced four published tools with an item structure.[31,32,33,37,38] Among the referenced tools, only the Psychometric Grading Framework was applied in its original form[33]; other tools were modified with[32] and without[31] justification by authors, or it was unclear how the tool was used because the assessment differed from the tool.[37,38]

Table 3: Approaches used to assess the adequacy of evidence-based practice measures across systematic reviews

Author (year)	Assessment approach described?	Validity, reliability, and other characteristics reported	Criteria of adequacy at the property level	Criteria of adequacy at the measure level	# of assessors	Tool referenced	Itemized structure	Modifications to the tool and justification
Albarqouni *et al.*[40] (2018)	Yes	Content validity, internal validity (internal consistency and dimensionality), responsive validity, discriminative validity, criterion validity, inter-rater reliability	"Established" if statistical test was significant	Shaneyfelt *et al.* method Level 1: established inter-rater reliability (if applicable), objective[b] and multiple ($\geq$3) types of established validity evidence	Unclear	NR	N/A	N/A
Belita *et al.*[36] (2020)	No explicit assessment conducted	Content, response process, internal structure, relationships to variables, internal consistency, inter-rater, test-re-test Other: acceptability (missing data, completion time)	The grouping of measures was hierarchical and based on the number of different sources of validity evidence. Four groups (eg, group 1 measures compiled four sources of validity evidence). Inclusion as "source of evidence" not described	NR	Two	The Standards of Psychological and Educational Testing,[41] 2014	Not item structured	N/A
Buchanan *et al.*[31] (2016)	Yes	Content validity, structural validity, hypothesis testing, cross-cultural validity, internal consistency, reliability Other: clinical utility	2-, 3-, or 4-point scale (scored between excellent, good, fair, poor); see Appendix IX	High-quality measures must score "excellent" for at least three measurement properties	One	COSMIN checklist (Terwee *et al.*,[42] 2012)	Yes	Modifications to the tool, without justification
Fernandez-Dominguez *et al.*[38] (2014)	Yes	Content validity, construct validity, criterion validity, floor/ceiling effects, responsiveness, internal consistency, reproducibility, and theoretical ground	3-point scale (clearly stated, partially stated, none)	Not conducted	Two	Terwee *et al.* quality criteria,[43] 2007	Yes	Modifications to the tool[c]
Glegg and Holsti[32] (2010)	Yes	Focus of the measure, scale construction, reliability, validity (includes rigor, content, construct, responsiveness)	Scale construction and reliability rated on a 3-point scale (excellent, adequate, poor); see Appendix IX	Overall utility is scored on a 3-point scale (excellent, adequate, poor) based on the rating of the two properties and on the validity evidence (although this was not scored)	One	CanChild Outcome Measures Rating Form and Guidelines,[44] 2004	Yes	Modifications to the tool, with justification

Table 3: (*Continued*)

Author (year)	Assessment approach described?	Validity, reliability, and other characteristics reported	Criteria of adequacy at the property level	Criteria of adequacy at the measure level	# of assessors	Tool referenced	Itemized structure	Modifications to the tool and justification
Kumaravel *et al.*[39] (2020)	Yes	Content validity, inter-rater reliability, internal validity, responsive validity, discriminative validity, construct validity, internal reliability, external validity	"Established" not described; assumed to be as in the Shaneyfelt *et al.*[34] paper - based on statistical significance	Shaneyfelt *et al.* method Level 1: established inter-rater reliability (if applicable), objective[b] and multiple (≥3) types of established validity evidence	Three	NR	N/A	N/A
Leung *et al.*[33] (2014)	Yes	Content validity, construct validity, criterion validity, internal consistency, test-retest reliability, and inter-rater reliability	4-point scale of strength (scored A-D); see Appendix IX	4-point scale of strength (good, adequate, weak, very weak) by combining the number and level of psychometric measures arising from Scale 1 See Appendix IX	Three	Psychometric Grading Framework (Leung *et al.*,[45] 2012)	Yes	No
Oude Rengerink *et al.*[37] (2013)	Yes	Reliability, validity content, structural, construct validity, responsiveness, internal consistency, test-retest reliability, and inter-rater reliability[a]	Yes/no checklist for when reliability and validity have been tested	"Valid and reliable tools" if both reliability and validity had been tested (not necessarily confirmed)	One (validated by second)	COSMIN checklist (Mokkink *et al.*,[46] 2010)	Yes	Modifications to the tool[c], without justification
Shaneyfelt *et al.*[34] (2006)	Yes	Content, internal structure (internal consistency and dimensionality), relationship to other variables (responsive, discriminative, criterion), inter-rater-Other: feasibility (ie, ease of implementation)	"Established" if statistical test was significant	L1: Established inter-rater reliability (if applicable), objective[b], multiple (≥3) types of established validity evidence (including evidence of discriminative validity) L2: Established inter-rater reliability (if applicable) L3: objective[b], non-self-reported	Two	The Standards of Psychological and Educational Testing,[47] 1999	Not item structured	N/A
Thomas and Kreptul[35] (2015)	No	Face, content, concurrent, construct validity; intra/inter-rater, internal, and test-retest reliability Other: item difficulty, item discrimination	Authors only report "all articles were assessed for validity and reliability"(p.108)	NR	NR	NR	N/A	N/A

COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; L1, Level 1: to distinguish between individuals; L2, Level 2: for evaluating program effectiveness; L3, Level 3: evaluation of evidence-based practice behaviors, N/A, not applicable; NR, not reported.
[a]These properties are reported in an additional file and grouped together as measurement properties.
[b]Objective (non-self-reported) outcome measure.
[c]The applied adequacy criteria diverge from the tool.

Systematic review recommendations

The Buchanan *et al.*[31] review was the only one to explicitly recommend specific EBP measures. Four other systematic reviews identified high-quality or adequate measures based on their criteria.[34,37,39,40] Table 4 presents an overview of the characteristics of the 27 recommended measures that were identified by these five systematic reviews.[48-92] The targeted populations of the five reviews were: all health care professionals (n = 3), occupational therapists (n = 1), and medical students (n = 1). Reviews occasionally presented diverging classifications by EBP domain or step for the same measure, while some did not provide classifications. Three other systematic reviews assessed the adequacy of measures; however, none of these reviews identified measures meeting their criteria for adequacy (see Table 3).[32,33,38]

Eight measures were deemed adequate by at least two systematic reviews (based on the criteria reported in Table 3). Providing a parameter of agreement between reviews would be arbitrary due to the heterogeneity in study aims, targeted population, and other inclusion criteria. Still, Albarqouni *et al.*[40] and Kumaravel *et al.*[39] identified the same six measures as adequate, even though the former focused on all health care professionals and the latter only on medical students. Four of those six measures were also judged adequate by Shaneyfelt *et al.*,[34] which covered all health care professionals, although most measures were developed for medical students and residents.

Summary of evidence

Table 5 presents a summary of the critical appraisal notations, methods used and reported, scope of reporting on psychometric findings, approach to assessing measures, and whether authors recommended specific measures across included systematic reviews.

Discussion

Quality of included systematic reviews

Although the JBI critical appraisal checklist results suggest better scores in more recent years (ie, notions are above 6.5 since 2018), our findings reveal important methodological shortcomings that affect the trustworthiness of findings.[17] As no systematic review included a critical appraisal of primary studies (Q5, Q6 of the JBI critical appraisal checklist), the extent to which primary studies are biased is unknown; in turn, authors may have underestimated or overestimated the adequacy of measures.[12,14] This is a major concern regarding risk of bias for all included reviews as the primary studies may have systematic flaws leading to inaccurate conclusions. It is vital for systematic reviews of psychometric properties to separately assess primary study quality and risk of bias in addition to assessing the adequacy of specific measures.[11,13,17,42,93,94] No review assessed the certainty of findings or addressed risk of bias, inconsistency, imprecision, or indirectness despite the availability of multiple tools for this purpose (eg, Grading of Recommendations Assessment, Development, and Evaluation approach).[11,23]

Additional major concerns of the included systematic reviews include i) the absence of an a priori protocol registration in 9/10 reviews, ii) low adherence to methodological and reporting guidelines in 5/9 reviews, and iii) failure to provide a detailed search syntax for at least one database in 5/10 reviews, all of which are considered best practices for systematic reviews.[17] Moreover, most search strategies were built without the expertise of an information specialist or librarian, despite this having been associated with superior review search quality.[95,96] Our findings are surprising given that 9/10 included reviews were published after 2009 when the PRISMA statement for the reporting of systematic reviews was introduced.[17] The COSMIN guidelines also existed for systematic reviews of measurement properties, first in 2010[97] and then in 2018.[11,23] The search strategy of one review was limited as the authors excluded all psychometric studies subsequent to the first measure development study and considered these studies as "duplicates"[39]; validity is a matter of continuous collection of evidence and subsequent psychometric studies on a measure are encouraged.[98,99]

All 10 systematic reviews present a high concern for risk of bias due to the absence of a quality and risk of bias assessment of primary studies; therefore, their recommendations and conclusions should be interpreted with caution. There is also a major concern for risk of bias for all but one review[36] as it was unclear, in the absence of a pre-published protocol, whether study eligibility criteria were prespecified or adapted post hoc. Important studies may have been missed due to the selection of studies by

Table 4: Characteristics of the 27 measures deemed adequate by five included systematic reviews

Measures	First author of the review	Studied populations	Description (as reported in corresponding reviews)	EBP domains				EBP steps				
				K	S	A	B	1	2	3	4	5
Fresno test (Ramos, 2003)[48]	Albarqouni[40]	Family practice residents and teachers, EBM experts (USA)[48]	Two clinical scenarios with 12 open-ended questions (fill in the blank and calculations), which are scored with standardized grading rubrics (four grading categories). Has been adapted for use in OT, PT, pharmacy, and in other languages (Brazilian, Portuguese). 40-60 min to complete, 10-15 min to mark.	x	x			x	x	x		
	Kumaravel[39]			x	x			x	x	x		
	Shaneyfelt[34]							x	x	x	x	
	Thomas and Kreptul[35]	Not deemed adequate by this review.										
Berlin Questionnaire (Fritsche *et al.*, 2002)[49]	Albarqouni[40]	EBM experts, medical students, and participants in EBP course (Germany)[49]; internal medicine residents[a] (USA)[50]	Two separate sets of 15 multiple-choice questions built around clinical scenarios and mainly focusing on epidemiological knowledge and skills. Has been translated in Dutch.	x	x			x	x	x		
	Kumaravel[39]			x	x					x		
	Shaneyfelt[34]							x	x	x	x	
MacRae *et al.* (2004)[51]	Albarqouni[40]	Surgery residents (Canada),[51] surgeons[b] [52]	Three journal articles, each followed by a series of short-answer questions and 7-point scales to rate the quality of elements of the study design.		x					x		
	Kumaravel[39]			x	x					x		
	Shaneyfelt[34]									x		
Taylor *et al.* (2001)[53]	Albarqouni[40]	Four groups of HCP with varying EBP expertise[53]; medical students in Norway[54] and Mexico[55]; various HCPs[56]; delegates at EBP conference[57]	Part I: Six multiple-choice questions, each with three items, with three potential answers, each requiring a true, false, or "don't know" response; the range of scores is -18 to $+18$.Part II: Seven statements related to the use of evidence in practice, and each scored using a 5-point Likert scale; the range of scores is 7 to 35.	x	x	x			x	x		
	Kumaravel[39]			x		x			x	x		
	Shaneyfelt[34]								x	x		
	Fernandez-Dominguez[38]	Not deemed adequate by this review.		x		x						
ACE tool (Ilic *et al.*, 2014)[58]	Albarqouni[40]	Medical students with varying EBP expertise (Australia)[58]	15 dichotomous-choice (yes or no) questions, based on a short patient scenario, a relevant search strategy, and a hypothetical article extract (scores ranged from 0 to 15). Items 1 and 2, asking the answerable question; items 3 and 4, searching literature; items 5-11 critical appraisal; items 12-15 relate to step 4 applying evidence to the patient scenario.	x	x			x	x	x		
	Kumaravel[39]			x	x			x	x	x	x	
Philibert *et al.* (2003)[59]	Buchanan[31]	OTs (USA)[59]	Self-report questionnaire containing four sections. Items included Likert-type scales with varying response options and two open-ended questions. Reading journals and sources used.			x	x					
	Oude Rengerink[37]								x			
	Glegg[32]	Not deemed adequate by this review.		x		x	x					
U-CEP (Kortekaas *et al.*, 2017)[60]	Albarqouni[40]	Postgraduate GP trainees, hospital trainees, GP supervisors, academic GPs, or clinical epidemiologists (Netherlands)[60]	Two formats: two sets of 25 comparable questions (six open-ended and 19 multiple-choice questions) and a combined set of 50 questions. Multiple-choice question scored 1 for correct and 0 for incorrect answer. Open-ended questions scored 0 to 3. Scores ranged from 0 to 33 for set A and 0 to 34 for set B.	x						x	x	
	Kumaravel[39]			x					x	x	x	

Table 4: (*Continued*)

Measures	First author of the review	Studied populations	Description (as reported in corresponding reviews)	EBP domains				EBP steps				
				K	S	A	B	1	2	3	4	5
EBPQ or revised Upton (Upton and Upton, 2006)[61]	Buchanan[31]	Allied health professions and health science services including OTs in the UK[62]	A 24-item self-report measure that assesses three subscales: knowledge, practice, and attitudes toward EBP. Knowledge/skills (14 items) are assessed collectively using a 7-point scale (1 =poor to 7=best). Practice is assessed with six items with a scale to determine the frequency with which that item has been completed over the past year on a 7-point scale ranging from never to frequently. Attitudes are assessed using four items also on a 7-point scale with higher scores indicating more positive attitudes towards EBP. Can be completed in 20 minutes.	x	x		x					
	Oude Rengerink[37]	Nurses (Italy)[63] translation in Italian by Romani						x	x	x	x	x
	Belita[36]	This measure was classified in group 4 (measures with one source of validity evidence).		x	x	x	x					
	Glegg[32]	Not deemed adequate by this review.		x	x		x					
	Leung[33]	Not deemed adequate by this review. (N.B. this review does not include “behavior” as an EBP domain.)		x	x	x						
Chiu (2009)[64,65]	Oude Rengerink[37]	MDs and nurses[64,65]	A self-report measure to assess EBP beliefs, attitudes, knowledge, skills, behaviors, and barriers. Respondents rate agreement on a 5-point Likert scale (from strongly agree to strongly disagree). EBP behaviors are defined by identifying the frequency of access to online databases.						x			
EBPP (McEvoy *et al.*, 2010)[66]	Oude Rengerink[37]	Multiple HCPs (Australia)[66]	Questionnaire on tracking down relevant evidence once a question is formulated and integrating research evidence with expertise.						x		x	
	Fernandez-Dominguez[38]	Not deemed adequate by this review.		x	x	x	x					
BARRIERS (Funk *et al.*, 1991)[67]	Buchanan[31]	Pediatric OTs (Australia, Taiwan, and UK[68];UK[69]; Australia[70]), dieticians, PTs, and speech therapists[71]	Four sub-scales and 28 items rated on a 4-point scale. Perceived barriers to research utilization (classified by study as “EBP behavior”).				x					
BART (Metcalfe *et al.*, 2001)[72]	Buchanan[31]	OTs (UK)[72]	Self-report questionnaire with two sections: perceived importance of research (7 questions) and perceived barriers (22 questions). Scores for sections 2 and 3 ranged from −7 to +7 and −22 to +22, respectively.			x	x					
	Fernandez-Dominguez[38]	Not deemed adequate by this review.				x						
	Glegg[32]	Not deemed adequate by this review.		x	x	x						
EROS (Pain *et al.*, 1996)[73]	Buchanan[31]	Pediatric OTs (Australia, Taiwan, and UK[68];UK[69]; Australia[70]), OTs (Canada)[74]	Self-report two-part questionnaire (self-rated knowledge of research concepts and participation in research and research orientation to practice) consisting of 38 items rated on a 5-point Likert scale.	x			x					
	Glegg[32]	Not deemed adequate by this review.		x			x					
KAB (Stronge and Cahill, 2012)[75]	Buchanan[31]	Final year OT students (Ireland)[75]	Self-report containing subjective and objective questions. Consists of four sub-scales: knowledge (5 items), attitudes (6 items), and future use of EBP (9 items) rated on a 6-point scale, and 17 additional questions on sources of evidence and demographics.	x		x	x					

Table 4: (*Continued*)

Measures	First author of the review	Studied populations	Description (as reported in corresponding reviews)	EBP domains				EBP steps				
				K	S	A	B	1	2	3	4	5
KAP (Van Mullem *et al.*, 1999)[76]	Buchanan[31]	Pediatric OTs (Australia, Taiwan, and UK[68]; UK[69]; Australia[70]), nurses and other HCPs, including OTs (USA)[77]	33-item self-report consisting of five factors; items rated on a 3-point scale and sub-scale scores determined. Knowledge of activities related to utilizing research; willingness to engage in activities related to utilizing research; ability to perform activities related to utilizing research.	x		x	x					
EBP and clinical effectiveness questionnaire (Upton and Lewis, 1998[78])	Buchanan[31]	Podiatrists, OTs, PTs, and speech therapists (UK)[79]	Five sections with varied response formats (visual analogue scales, semantic differentials, Likert-type scales) and a section for open comments. Perceived knowledge of EBP and its individual steps; attitudes to EBP; frequency of completing EBP steps; barriers and solutions to implementing EBP.	x		x	x					
	Leung[33]	Not deemed adequate by this review. (N.B. this review did not include "behavior" as an EBP domain).		x	x	x						
	Fernández-Domínguez[38]	Not deemed adequate by this review.		x	x	x	x					
Jette (2003)[80]	Oude Rengerink[37]	PTs[80]	Frequency of reading articles, use of databases, literature, and access to practice guidelines).						x			
	Fernández-Domínguez[38]	Not deemed adequate by this review.		x		x	x					
KACE (Hendricson, 2011)[81]	Oude Rengerink[37]	Dental faculty[81]	KACE: Evidence Based Practice Knowledge, Attitudes, Access and Confidence Evaluation. Accessing evidence: frequency of accessing the Cochrane Library.						x			
Cobban (2007)[82]	Oude Rengerink[37]	Dental hygienists[82]	Research utilization								x	
Veeramah (2004)[83]	Oude Rengerink[37]	Nurses and midwives[83]	Frequency of using research findings in practice								x	
Chernick (2010)[84]	Oude Rengerink[37]	MDs, pediatric interns and residents, EBP experts[84]	Self-reported practice of EBP: frequency of searching articles to answer clinical questions, generation of clinical questions applicable to patients' diagnostic or therapeutic plan.					x	x			
Filippini (2011)[85]	Oude Rengerink[37]	Nurses[85]	Frequency of reading guidelines and scientific journals; having modified practice last year; frequency of using EBP.						x		x	
Kahveci and Meads (2009)[86]	Oude Rengerink[37]	Primary care physicians[86]	Use of resources and percentage of EBP.						x		x	
Lavis *et al.* (2010)[87]	Oude Rengerink[37]	GPs, specialists, nurses, health workers[87]	Access to evidence and use of sources; use of evidence and change in practice attributed to particular sources of research evidence.						x		x	

Table 4: (*Continued*)

Measures	First author of the review	Studied populations	Description (as reported in corresponding reviews)	EBP domains				EBP steps				
				K	S	A	B	1	2	3	4	5
Bennett *et al.* (1987)[88]	Shaneyfelt[34]	Medical students in clerkships[88]	Open-ended free-text questions, fill-in-the-blank questions, and calculations relating to two pediatric clinical scenarios; scored using a standardized grading rubric that includes examples of acceptable answers and specifies 4 or 5 grading categories (not evident, minimal, and/or limited, strong, excellent), each of which is associated with a point value.							x		
Weberschock *et al.* (2005)[89]	Shaneyfelt[34]	Third-year medical students and students with advanced training in EBM[89]	Five sets of 20 multiple-choice questions (5 "easy," 10 "average," and 5 "difficult") linked to clinical scenarios and pertaining to data from published research articles.	x	x							
Haynes *et al.* (1990,[90] 1993[91]); McKibbon *et al.* (1990)[92]	Shaneyfelt[34]	Physicians and physicians in training[91]; novice and expert clinicians and librarians[90,92]	Search output; relative recall; precision; article "relevance" (library computer based).				x		x			

ACE, assessing competency in EBM; BART, barriers and attitudes to research in the therapies; EBM, evidence-based medicine; EBP, evidence-based practice; EBPP, evidence-based practice profile; EBPQ, evidence-based practice questionnaire; EROS, Edmonton research orientation survey; GP, general practitioner; HCP, health care professional; KAB, knowledge, attitudes, and behavior; U-CEP, the Utrecht Questionnaire; MD, medical doctor; OT, occupational therapy(ist); PT: physical therapy(ist)
EBP domains: K, knowledge; S, skills; A, attitudes; B, behaviors
EBP steps: 1, ask; 2, acquire; 3, appraise; 4, apply; 5, assess
■ = the systematic review did not report this categorization (ie, EBP domains or steps)
[a]reported by Albarqouni and Shaneyfelt only
[b]reported by Shaneyfelt only

Table 5: Summary of evidence of the included systematic reviews

Authors (year)	JBI critical appraisal checklist for systematic reviews and research syntheses	Methods	Reporting of psychometric findings	Approach to assessing measures	Did the authors recommend measures (Yes/No)
Albarqouni *et al.*[40] (2018)	7	• PRISMA flowchart included • PRISMA checklist included in additional file • Search conducted by a senior information specialist • Reference checking was performed • Reasons for exclusion of articles reported	• For a subset of higher-quality measures only • No scores provided • Evidence compiled by measure	• Shaneyfelt *et al.* method (level 1 only) • Number of assessors unclear	Yes, through the assessment and reporting of adequate measures
Belita *et al.*[36] (2020)	7	• Protocol registered • Cochrane methodology mentioned in abstract only • PRISMA reporting guidelines • PRISMA flowchart included • Two independent reviewers for study eligibility • Detailed search syntax available • Search conducted by a health sciences librarian • Reasons for exclusion of articles reported	• For all measures • Scores provided for all included properties • Evidence compiled by measure, then by study • Categorization by practice setting	• The grouping of measures based on sources of validity evidence was conducted by two assessors • Based on The Standards 2014[41]	Not explicitly, although measures are categorized by number of sources of validity
Buchanan *et al.*[31] (2016)	5.5	• Cochrane approach mentioned for "databases searched" in methods • PRISMA flowchart included • Two independent reviewers for study eligibility • Reference checking was performed • Reasons for exclusion of articles reported	• For all measures • Scores provided for some properties • Evidence compiled by study • Evidence deemed the same for adaptations of measures	• Modified and applied tool: COSMIN checklist[42] • One assessor	Yes, explicitly recommended measures
Fernaández-Domínguez *et al.*[38] (2014)	6	• PRISMA flowchart included • Two independent reviewers for study eligibility • Detailed search syntax available • Reference checking was performed • Reasons for exclusion of articles reported	• For all measures • Scores provided for some properties • Evidence compiled by study	• Modified and applied tool: Terwee checklist[43] • Two assessors	No measures were deemed adequate after assessment
Glegg and Holsti[32] (2010)	3	• Reference checking was performed	• For all measures • Scores provided for some properties • Evidence compiled by measure • Evidence deemed the same for adaptations of measures	• Modified and applied tool: CanChild Outcome Measures Rating Form[44] • One assessor	No measures were deemed adequate after assessment

Table 5: (*Continued*)

Authors (year)	JBI critical appraisal checklist for systematic reviews and research syntheses	Methods	Reporting of psychometric findings	Approach to assessing measures	Did the authors recommend measures (Yes/No)
Kumaravel *et al.*[39] (2020)	6.5	• PRISMA reporting guidelines mentioned in abstract only • PRISMA flowchart included • PRISMA checklist included in additional file • Two independent reviewers only at full-text level of study eligibility • Detailed search syntax available • Search conducted by an information specialist • Reference checking was performed • Reasons for exclusion of articles reported	• For all measures • Scores provided for some properties • Evidence compiled by measure	• Shaneyfelt *et al.* method (level 1 only) • Three assessors	Yes, through the assessment and reporting of adequate measures
Leung *et al.*[33] (2014)	7.5	• PRISMA reporting guidelines • PRISMA flowchart included • Two independent reviewers only at full-text level of study eligibility • Detailed search syntax available • Reasons for exclusion of articles reported	• For all measures • Scores provided for some properties • Evidence compiled by study	• Tool: Psychometric Grading Framework[45] • Three assessors	No measures were deemed adequate after assessment
Oude Renger-ink *et al.*[37] (2013)	5	• PRISMA reporting guidelines • PRISMA flowchart included • Two independent reviewers for study eligibility • Detailed search syntax available • Reasons for exclusion of articles reported	• For all measures • Scores provided for some properties (in additional file) • Evidence compiled by study • Evidence deemed the same for adapta tions of measures	• Modified and applied tool: COSMIN checklist[46] • One assessor, validated by second	Yes, through the assessment and reporting of adequate measures
Shaneyfelt *et al.*[34] (2006)	7	• PRISMA flowchart included • Two independent reviewers for study eligibility • Reference checking was performed • Reasons for exclusion of articles reported	• For a subset of higher-quality measures only • Scores provided for some properties • Evidence compiled by measure	• Self-developed method • Based on The Standards 1999[47] • Two assessors	Yes, through the assessment and reporting of adequate measures
Thomas and Kreptul[35] (2015)	2	• PRISMA flowchart included • Two independent reviewers for study eligibility • Reasons for exclusion of articles reported (number of studies per reason unknown)	• For all measures • Scores provided for some properties • Evidence compiled by study	No assessment described	Not explicitly, although one measure is emphasized over others

COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; The Standards, The Standards for Educational and Psychological Testing only one reviewer at the abstract[32,33,39,40] or full-text[32,40] level. One review[32] may have also missed important studies due to restrictions in the eligibility criteria that were imposed without rationale. In two reviews,[31,32] some bias may have been introduced through the data collection process, as extraction was completed by one reviewer without mention of methods to minimize errors in data extraction, such as piloting or verification by another individual. One review[35] had a particularly high risk of bias due to inadequacies in methodology and reporting, such as failure to report the eligibility criteria, the search strategy, and the methods used to assess the adequacy of included measures.

Our results show limited reporting of psychometric evidence regarding the primary study design and methods, sample size, participants, statistical analyses, and results. The absence of in-depth information hinders end users from interpreting the data and corroborating the authors' conclusions.[100] Many reviews did not report actual scores of psychometric tests; rather, they reported broad conclusions on the strength of properties (eg, "[property] demonstrated," "evidence of [property]," or "good [property]").[31] Specific to construct validity, most authors did not provide a priori hypotheses or the direction and magnitude of the correlation. No review reported on the measurement model (ie, reflective or formative), which has implications for interpreting internal consistency scores.[46]

Many authors merely stated that content validity had been established, yet evaluating content validity requires a judgment on item relevance, comprehensiveness, comprehensibility, and quality.[101] In some cases, authors summarized the item review method using keywords (eg, "expert panel") but failed to describe the concepts to be measured, the theoretical basis, or conceptual framework related to the included domains and the quality of the development process.[101] Four reviews did not report the intended purpose of measures, which may thwart attempts to interpret the psychometric evidence and determine the adequacy of measures for use.

The reporting on characteristics of the testing samples (eg, profession, country, practice setting, level of expertise) was often insufficient to allow end users to judge whether the individuals within their setting are comparable.[94] In some cases, the psychometric evidence for a measure was compiled without distinguishing between the different testing samples; the link between testing sample and psychometric scores must be explicit as differences in sample characteristics may lead to variations in the validity of inferences.[11,102] In three reviews,[31,32,37] the psychometric evidence associated with a measure was assumed to be the same for its modified version; however, each version of a measure must be considered separately.[11]

End users may need to make decisions on the appropriateness of a measure based on a trade-off between psychometric strength and pragmatic considerations.[103] To do so, there is a need to improve the reporting of feasibility and acceptability characteristics of measures as this was only addressed by three reviews.[31,34,36] No review discussed the accessibility of measures, despite the 1999 version of the Standards of Psychological and Educational Testing (hereafter referred to as "The Standards") stressing the importance of ensuring accommodation andfairnessintesting.[41] While these aspects are generally highlighted for patient-reported outcome measures (PROMs), they should also be considered when assessing EBP in clinical and educational contexts.

Superficial reporting of psychometric evidence could be attributed to the far-reaching objectives of most reviews regarding the inclusion of target populations, types of measures, EBP domains, and psychometric properties. It is recommended to conduct an in-depth analysis of each distinct psychometric property for each measure.[11,23,43] The amount of work required to meet such broad objectives may have been unmanageable. The review by Belita *et al.*[36] provides the most detailed, comprehensive, and useful reporting of psychometric evidence. Future studies are needed that narrow the aim to one or a couple of properties, or to specific measures to permit an exhaustive reporting and analysis of psychometric evidence.

Approaches used to assess the adequacy of evidence-based practice measures

Authors utilized varying criteria to assess the adequacy of measures a finding which is consistent with other reviews.[12,14] The diversity of approaches may hamper the comparison of findings between reviews and create confusion for end users looking for best practices in assessing the adequacy of EBP measures.

Although two reviews[35,36] did not report an assessment approach the results and conclusions of these reviews suggest an implicit assessment or ranking of measures by the authors.

The four itemized tools that were included in five reviews had diverse forms regarding the included measurement properties number of items per property, and rating scale. Most authors modified the tool by omitting measurement properties and reducing the number of items without justification, suggesting potentially biased findings.

Glegg and Holsti[32] were the only ones to justify the modifications; the rationale on why they did not evaluate the strength of validity evidence is questionable. These authors applied a modified version of the CanChild Form and Guideline,[44] which was developed for PROMs with items relating to the International Classification of Functioning, Disability and Health (ICF) framework; in our view, this is impractical for use with EBP measures as the focus of measurement is unrelated to body functions, body structures, or participation. Oude Rengerink *et al.*[37] and Buchanan *et al.*[31] referenced the COSMIN checklist as a tool to assess the quality of psychometric properties, when in fact, it is meant to assess the methodological quality of studies on psychometric properties. The misuse of the COSMIN checklist has been discussed elsewhere and can lead to misconstrued conclusions.[46] Oude Rengerink *et al.*[37] and Fernáindez-Domí-nguez *et al.*[38] presented inconsistencies between the methods and results by disregarding the referenced tool and determining the adequacy of measures based on the availability of reporting of psychometric tests in the literature. It is important to consider the actual psychometric study results to provide substantiated conclusions on the adequacy of measures. Among the reviews that referenced itemized tools, Leung *et al.*[33] stood out as having the most compelling approach through the use of The Psychometric Grading Framework. However, given that the results of psychometric properties (ie, scale 1) and intended uses of measures were not reported, it is challenging for end users to judge the adequacy of psychometric properties for specific contexts of assessment.[43] For instance, reliability is more pertinent when using a questionnaire for discerning between people and responsiveness, and is more relevant when evaluating an intervention effect.[43]

The Standards, which were referred to by Shaneyfelt *et al.*[34] and Belita *et al.*,[36] do not report an item structure or rating scale; this may explain why Belita *et al.* did not report an explicit assessment and why Shaneyfelt *et al.* developed their own criteria. The Shaneyfelt *et al.* criteria, which was applied by three reviews,[34,39,40] and the inherent grouping of measures by Belita *et al.* reflect a judgment of adequacy based on the overall number of sources of validity evidence.

The Standards articulate five categories of validity evidence (test content, response process, internal structure, relations to other variables, consequences). These four reviews included "test content," "internal structure," and "relations to other variables," and only Belita *et al.* included "response process." This aligns with findings from a study that showed that "consequences" and "response process" received little attention and low ratings for clinical teaching assessments.[104] While "consequences" is the most controversial source of validity evidence (as the consequences of tests have no part in the validity of inferences), the consequences of tests, especially in the context of high-stakes clinical or educational assessments, must be discussed.[105]

Shaneyfelt *et al.*[34] were the only authors to assess the strength of psychometric evidence while considering the purpose of the assessment (ie, levels 1 to 3); this is an important consideration as the needs of end users will vary, and an assessment of the adequacy of measures must be tailored to the intended use. The authors distinguished adequate measures for discriminating between individuals with different EBP expertise, evaluating a group effect and evaluating EBP behavior. The two reviews that adopted the Shaneyfelt *et al.* method[39,40] did not discuss the purpose of assessment and only used level 1 criteria. The Shaneyfelt *et al.* criteria state that measures must be "objective (non-self-report)" to be classified into levels 1 to 3. While we agree that self-report instruments may incur social desirability or recall bias, end users may wish to consider self-report measures given their practicality of administration.[106,107]

Further, this approach uses the term "established" to denote statistically significant results for psychometric tests and inclusion as a source of validity. The term "established" has a definitive connotation that clashes with the integrative and ongoing nature of validity as per The Standards, whereby validity is understood as a matter of degree rather than all-or-none.[41,108] Also, while the topic of retiring statistical significance has been discussed elsewhere,[109] there are many reasons for non-significant findings, including a limited sample size, missing data, or inappropriate analysis.[43] Although these precisions should be reported and interpreted by authors of systematic reviews, the term "established" to interpret psychometric evidence can be misleading.

Recommendations for evidence-based practice measures

The failure by all but one review[31] to provide recommendations for the use of EBP measures is surprising. Providing clear recommendations for the use of measures is the primary aim of psychometric systematic reviews seeking to support end users in selecting robust and relevant measures.[94] The need to recommend measures is also highlighted in a recent study where 49% of systematic reviews on health-related outcome measures provided a recommendation for at least one measure.[13]

Still, we identified 27 measures that were deemed adequate for use by five of the 10 systematic reviews. These results must be interpreted with caution due to the diversity of approaches used to assess the adequacy of measures and often questionable methodological quality of included reviews. The exploratory mapping revealed an uneven distribution of these measures across steps and domains, with step 2 (accessing evidence) being the most represented step (n = 16) and knowledge the most represented EBP domain (n = 17). Only one measure was associated with step 5 (assess outcomes), which mirrors the paucity of literature examining the impact of EBP interventions or competencies on changes in care processes or patient outcomes.[110]

Given the heterogeneity of review aims and methods, it would be arbitrary to provide a parameter of agreement across reviews on the adequacy of measures. As such, we urge readers to cautiously interpret Table 4, as most of the corroborated assessment results stemmed from three reviews that applied the Shaneyfelt *et al.* criteria. We also conclude that it would not be wise to suggest a "best" EBP measure. Rather, we urge readers to use the methodological and theoretical concepts discussed in this review as a starting point in conducting an appraisal process to select an appropriate measure relevant to their assessment context.

Strengths and weaknesses of the review

This is the first synthesis of systematic reviews on EBP measures in health care. We attempted to unify this complex topic using robust methods, adhering to JBI guidelines and PRISMA, with a research team that includes complementary methodological strengths and content expertise.

We recognize that there is an inherent tension throughout the paper regarding the conceptualization of validity and associated taxonomies. Certain statements or methodologies may suggest adherence to a specific view of validity and may seem contradictory at times to the reader. Whenever possible, we opted to keep the integrity of the statements used in the included systematic reviews when describing findings, as it was not our goal to favor one conceptualization of validity.

We also recognize certain limitations to our review. As our search was limited to systematic reviews on EBP measures, we may have excluded reviews on implementation of best practices or research utilization outcomes that may include EBP measures. We opted for a descriptive approach rather than an evaluative approach (ie, with a scoring system) as no current guidelines exist to substantiate such methods.

Also, psychometric terminology varied across reviews, which complicated the task of mapping the reporting of properties to the COSMIN taxonomy.[23] We selected this taxonomy as it offered the most discriminative classification (ie, 11 distinct characteristics). In Appendix VIII, we added bold case letters when the terminology differed and an additional row if the property did not fit within our pre-defined taxonomy. For the JBI checklist for critical appraisal results, the numerical scores serve to ease interpretation by allowing a rank ordering of reviews; however, these scores must be interpreted with caution as the relative importance of certain quality indicators over others is unknown.

Due to missing data on included measures in three reviews,[34,37,40] the total number of identified measures (ie, 204) across the included reviews is an understatement. Attempts at gaining the names of all included measures were unsuccessful after correspondence with the authors due to data destruction,[34] unavailable list,[40] or non-response by authors.[37]

Another limitation of this umbrella review lies in the inherited limitations of included systematic reviews, which did not assess primary study quality and rarely adhered to methodological and reporting guidelines. As the categorization of recommended measures by EBP step or domain differed across systematic reviews for the same measure, we have chosen to present all classifications reported by included systematic reviews (when applicable) to allow for interpretation by the reader.

Strengths and weaknesses in relation to existing literature

Generally, this study produced results that are consistent with studies appraising systematic reviews on health-related outcome measures and on quality-of-life instruments, although our findings on the appraisal of primary study quality are poorer (ie, 0% compared to 17%,[14] 30%,[13] and 41%.[12] We agree with the conclusions of these studies that there is still an urgent need for improvement regarding the search strategy, quality assessment of included studies, quality assessment of instruments, and data synthesis in psychometric systematic reviews. Developing an extension to the PRISMA checklist for systematic reviews of measures could improve reporting and reduce research wastage.[111]

Our analysis reveals a tension in how validity is conceptualized in included reviews, which mirrors an intricate and longstanding discussion in the literature.[105,112,113] This tension involves an approach to validity where the interpretation of scores on statistical tests establishes the strength of a measure, and another where validity is a unified concept and various sources of evidence support the inferences made by a measure.[9,41,99,112] These discourses are apparent in the language and methods used by authors.

Eight reviews[31-33,35,37-40] refer to "the validity and reliability of a measure," which demonstrates that validity is viewed as a property of the measure, while two reviews[34,36] view validity as a property of inferences and interpretations.[114] In terms of methods, nine reviews consider the scores on statistical tests for separate kinds of validity (eg, construct, content, criterion) to assess the adequacy of mea-sures.[31-35,37-40] However, three of these reviews[34,39,40] also report cumulating "sources of evidence" to support the validity of findings, revealing a conceptual discrepancy. Although the concept of validity as a property of the measure is refuted by some,[105,108,115] the fact that the majority of reviews conceptualize validity as such may suggest that it serves a pragmatic need.[99]

An international consensus statement was published 10 years ago to guide the classification and development of EBP assessment tools, which introduced the Classification Rubric for EBP Assessment Tools in Education (CREATE) framework.[8] While this statement highlights principles to be considered during tool development, it does not discuss the psychometric standards for EBP measures. There is no empirical evidence supporting the criterion of adequacy for the broad range of EBP assessment types (ie, self-report to activity monitoring) or domains (ie, knowledge, skills, attitudes, behaviors). While COSMIN offers detailed guidance on conducting systematic reviews on measurement properties of PROMs, it is unclear to what extent this guidance can be applied to the range of EBP measures.[11,23]

Conclusions and recommendations

Recommendations for practice

Results from this review should be considered before any future reviews on EBP measures are conducted to avoid similar methodological shortcomings and minimize the risk of bias. These findings suggest that assessing the adequacy of EBP measures is complex and requires an advanced understanding of psychometric theory and methods. We encourage greater collaboration between EBP researchers, educators, and psychometricians, a conclusion that has been suggested by other investigators.[12,13,103] In selecting an appropriate measure, we emphasize the careful trade-off between psychometric rigor, pragmatic aspects, and the applicability to the aims of an assessment. Before evaluating the psychometric evidence, end users should start by ensuring that the content validity is adequate concerning the assessment context, and the examination of content validity comprises more than the measure development process and requires a thorough evaluation by end users.[101,104,116] We also reinforce the ongoing nature of validity, as new findings may refute what is being claimed in included systematic reviews.

Recommendations for research

There is a need to determine (through consensus-based methods) the most appropriate application of validity theory and methodology as it pertains to EBP assessments. This process may include an analysis of the theoretical tensions regarding validity, and of the applicability of psychometric tools and standards to different types of EBP assessments. We also recognize the need for a practical resource for EBP researchers and educators to apply criteria of adequacy and select an EBP measure for a specific use. This resource could consist of a network of theoretical and psychometric evidence that supports the use of specific measures for a range of assessment purposes in a particular context. Finally, methodological guidance is required for umbrella reviews of systematic reviews of measurement properties and for assessing the certainty of findings in umbrella reviews.

Given the rapid development of EBP measures in the past two decades, this umbrella review compares and synthesizes the systematic review literature and provides timely conceptual and methodological direction to future studies seeking to advance the theory and practice of measuring EBP.

Acknowledgments

Nancy Mayo, PhD, McGill University, for providing feedback on the draft manuscript.

This study partially contributes towards the doctoral research of JRD. JRD is supported by a doctoral fellowship from the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreial (CRIR) and from the Richard and Edith Strauss Foundation.

Funding

This study was funded by the Canadian Institutes of Health Research (grant number 148544). The funding body did not influence the study results and was not involved in the study in any way.

Availability of data

Data are available on reasonable request to the authors.

References

1. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet 2017;390 (10092):415–23.
2. Guyatt GH. Evidence-based medicine. ACP J Club 1991;114 (2.):A-16.
3. Melnyk BM, Fineout-Overholt E, Gallagher-Ford L, Kaplan L. The state of evidence-based practice in US nurses: critical implications for nurse leaders and educators. JONA J Nurs Adm 2012;42(9):410–7.
4. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996:71–2.
5. Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayian K, *et al*. Sicily statement on evidence-based practice. BMC Med Educ 2005;5(1):1–7.
6. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. BMJ 1995;310 (6987):1122–6.
7. Satterfield JM, Spring B, Brownson RC, Mullen EJ, Newhouse RP, Walker BB, *et al*. Toward a transdisciplinary model of evidence-based practice. Milbank Q 2009;87 (2):368–90.
8. Tilson JK, Kaplan SL, Harris JL, Hutchinson A, Ilic D, Niederman R, *et al*. Sicily statement on classification and development of evidence-based practice learning assessment tools. BMC Med Educ 2011;11(1):1–10.
9. Downing SM. Validity: on the meaningful interpretation of assessment data. Med Educ 2003;37(9):830–7.
10. Crossley J, Humphris G, Jolly B. Assessing health professionals. Med Educ 2002;36(9):800–4.
11. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27(5):1147–57.
12. Terwee CB, Prinsen CAC, Garotti MR, Suman A, De Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res 2016;25(4):767–79.
13. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, *et al*. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res 2009;18(3):313–33.
14. Lorente S, Viladrich C, Vives J, Losilla J-M. Tools to assess the measurement properties of quality of life instruments: a meta-review. BMJ Open 2020;10(8):e036038.
15. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and. PLoS One 2012;7(11):e49667.
16. Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Chapter 10: Umbrella Reviews. In: Aromataris, E, Munn, Z, editors. JBI Manual for Evidence Synthesis [internet]. Adelaide: JBI; 2020 [cited 2020 Oct 29]. Available from: https://synthesismanual.jbi.global.
17. Moher D, Liberati A, Tetzlaff J, Altman. DG; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336–41.
18. CADTH. Strings attached: CADTH's database search filters [internet]. Ottawa: CADTH; 2014 [cited 2021 Feb 18].

Available from: https://www.cadth.ca/resources/finding-evidence/strings-attached-cadths-database-search-filters.
19. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015;13(3):132–40.
20. Shrier I, Boivin J-F, Steele RJ, Platt RW, Furlan A, Kakuma R, *et al*. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol 2007;166(10):1203–9.
21. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health 2018;21(3):95–100.
22. Aromataris E, Fernandez RS, Godfrey C, Holly C, Khalil H, Tungpunkom P. Methodology for JBI umbrella reviews. Joanna Briggs Institute Reviewers' Manual. Adelaide; JBI. 2014 [cited 2020 Oct 29]. Faculty of Science, Medicine and Health - Papers: part A. 3344. Available from: https://ro.uow.edu.au/smhpapers/3344.
23. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, *et al*. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) -user manual. Qual Life Res 2018;27(5):1147–57.
24. Cambridge English Dictionary. Recommendation. In: Cambridge English Dictionary [internet]. n.d. [cited 2020 Dec 11]. Available from: https://dictionary.cambridge.org/dic-tionary/english/recommendation.
25. Pieper D, Antoine S-L, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 2014;67(4):368–75.
26. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. *et al.* Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme [internet]. 2006 [cited 2021 Feb 1]. Available from: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf.
27. Tricco AC, Soobiah C, Antony J, Cogo E, Mac Donald H, Lillie E, *et al*. A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. J Clin Epidemiol 2016;73:19–28.
28. Altman DG. Practical statistics for medical research. CRC Press; 1990.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
30. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.
31. Buchanan H, Siegfried N, Jelsma J. Survey instruments for knowledge, skills, attitudes and behaviour related to evidence-based practice in occupational therapy: a systematic review: evidence-based practice survey instruments. Occup Ther Int 2016;23(2):59–90.
32. Glegg SMN, Holsti L. Measures of knowledge and skills for evidence-based practice: a systematic review. Can J Occup Ther Ott 2010;77(4):219–32.
33. Leung K, Trevena L, Waters D. Systematic review of instruments for measuring nurses' knowledge, skills and attitudes for evidence-based practice. J Adv Nurs 2014;70(10):2181–95.
34. Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, *et al*. Instruments for evaluating education in evidence-based practice: a systematic review. JAMA 2006;296(9):1116–27.
35. Thomas RE, Kreptul D. Systematic review of evidence-based medicine tests for family physician residents. Fam Med 2015;47(2):101–17.
36. Belita E, Squires JE, Yost J, Ganann R, Burnett T, Dobbins M. Measures of evidence-informed decision-making competence attributes: a psychometric systematic review. BMC Nurs 2020;19(1):44.
37. Oude Rengerink K, Zwolsman SE, Ubbink DT, Mol BWJ, van Dijk N, Vermeulen H. Tools to assess evidence-based practice behaviour among healthcare professionals. Evid Based Med 2013;18(4):129–38.
38. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gomez JE. Validity and reliability of instruments aimed at measuring evidence-based practice in physical therapy: a systematic review of the literature: evidence-based physiotherapy instruments. J Eval Clin Pract 2014;20(6):767–78.
39. Kumaravel B, Hearn JH, Jahangiri L, Pollard R, Stocker CJ, Nunan D. A systematic review and taxonomy of tools for evaluating evidence-based medicine teaching in medical education. Syst Rev 2020;9:1–12.
40. Albarqouni L, Hoffmann T, Glasziou P. Evidence-based practice educational intervention studies: a systematic review of what is taught and how it is measured.
41. American Educational Research Association, American Psychological Association. National Council on Measurement Education. The Standards for Educational and Psychological Testing. American Educational Research Association; 2014.
42. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21(4):651–7.
43. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42.
44. Law, M. Outcome measures rating form. A product from CanChild Centre for Childhood Disability Research

[internet]. 2004 [cited 2021 Feb 01]. Available from: https://www.canchild.ca/en/resources/137-critical-review-forms-and-guidelines.
45. Leung K, Trevena L, Waters D. Development of an appraisal tool to evaluate strength of an instrument or outcome measure. Nurse Researcher 2012;20(2).
46. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10(1):22.
47. American Educational Research Association, American Psychological Association. National Council on Measurement in Education. The Standards for Educational and Psychological Testing. American Educational Research Association; 1999.
48. Ramos KD. Validation of the Fresno test of competence in evidence based medicine. BMJ 2003;326(7384):319–21.
49. Fritsche L. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. BMJ 2002;325(7376):1338–41.
50. Akl EA, Izuchukwu IS, El-Dika S, Fritsche L, Kunz R, Schünemann HJ. Integrating an evidence-based medicine rotation into an internal medicine residency program. Acad Med 2004;79(9):897–904.
51. MacRae HM, Regehr G, Brenneman F, McKenzie M, McLeod RS. Assessment of critical appraisal skills. Am J Surg 2004;187(1):120–3.
52. MacRae HM, Regehr G, McKenzie M, Henteleff H, Taylor M, Barkun J, *et al*. Teaching practicing surgeons critical appraisal skills with an internet-based journal club: a randomized, controlled trial. Surgery 2004;136(3):641–6.
53. Taylor R, Reeves B, Mears R, Keast J, Binns S, Ewings P, *et al*. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. Med Educ 2001;35(6):544–7.
54. Bradley P, Oterholt C, Herrin J, Nordheim L, Bjørndal A. Comparison of directed and self-directed learning in evidence-based medicine: a randomised controlled trial. Med Educ 2005;39(10):1027–35.
55. Sanchez-Mendiola M, Kieffer-Escobar LF, Marín-Beltrán S, Downing SM, Schwartz A. Teaching of evidence-based medicine to medical students in Mexico: a randomized controlled trial. BMC Med Educ 2012;12(1):1–14.
56. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial [ISRCTN46272378]. BMC Med Educ 2004;4(1):30.
57. Bradley P, Herrin J. Development and validation of an instrument to measure knowledge of evidence-based practice and searching skills. Med Educ Online 2004;9(1):4354.
58. Ilic D, Nordin RB, Glasziou P, Tilson JK, Villanueva E. Development and validation of the ACE tool: assessing medical trainees' competency in evidence based medicine. BMC Med Educ 2014;14(1):114.
59. Philibert DB, Snyder P, Judd D, Windsor M-M. Practitioners' reading patterns, attitudes, and use ofresearch reported in occupational therapy journals. Am J Occup Ther 2003;57(4):450–8.
60. Kortekaas MF, Bartelink M-LE, de Groot E, Korving H, de Wit NJ, Grobbee DE, *et al*. The Utrecht questionnaire (U-CEP) measuring knowledge on clinical epidemiology proved to be valid. J Clin Epidemiol 2017;82:119–27.
61. Upton D, Upton P. Development of an evidence-based practice questionnaire for nurses. J Adv Nurs 2006;53(4):454–8.
62. Upton D, Upton P. Knowledge and use ofevidence-based practice by allied health and health science professionals in the United Kingdom. J Allied Health 2006;35(3):127–33.
63. Romani S, Forni C. Know-how, attitudes and obstacles to the implementation ofevidence-based practice: a descriptive study conducted within Azienda USL di Reggio Emilia. L'Infermiere 2010;47(4):23–8.
64. Chiu Y-W, Weng Y-H, Lo H-L, Hsu C-C, Shih Y-H, Kuo KN. Comparison of evidence-based practice between physicians and nurses: a national survey of regional hospitals in Taiwan. J Contin Educ Health Prof 2010;30(2):132–8.
65. Chiu Y-W, Weng Y-H, Lo H-L, Ting H-W, Hsu C-C, Shih Y-H, *et al*. Physicians' characteristics in the usage of online database: a representative nationwide survey of regional hospitals in Taiwan. Inform Health Soc Care 2009;34(3):127–35.
66. McEvoy MP, Williams MT, Olds TS. Development and psychometric testing of a trans-professional evidence-based practice profile questionnaire. Med Teach 2010;32(9):e373–80.
67. Funk SG, Champagne MT, Wiese RA, Tornquist EM. BARRIERS: the barriers to research utilization scale. Appl Nurs Res 1991;4(1):39–45.
68. Brown T, Tseng MH, Casey J, McDonald R, Lyons C. Knowledge, attitudes, practices and barriers of pediatric occupational therapists to evidence-based practice and research utilization. World Fed Occup Ther Bull 2009;60(1):38–48.
69. Lyons C, Casey J, Brown T, Tseng M, McDonald R. Research knowledge, attitudes, practices and barriers among paediatric occupational therapists in the United Kingdom. Br J Occup Ther 2010;73(5):200–9.
70. Lyons C, Brown T, Tseng MH, Casey J, McDonald R. Evidence-based practice and research utilisation: perceived research knowledge, attitudes, practices and barriers among Australian paediatric occupational therapists. Aust Occup Ther J 2011;58(3):178–86.

71. Jose Closs S, Lewin BJ. Perceived barriers to research utilization: a survey of four therapies. Br J Ther Rehabil 1998;5(3):151–5.
72. Metcalfe C, Lewin R, Wisher S, Perry S, Bannigan K, Moffett JK. Barriers to implementing the evidence base in four NHS therapies: dietitians, occupational therapists, physiotherapists, speech and language therapists. Physiotherapy 2001;87(8):433–41.
73. Pain K, Hagler P, Warren S. Development ofan instrument to evaluate the research orientation of clinical professionals. Can J Rehabil 1996;9:93–100.
74. Pain K, Magill-Evans J, Darrah J, Hagler P, Warren S. Effects of profession and facility type on research utilization by rehabilitation professionals. J Allied Health 2004;33(1):3–9.
75. Stronge M, Cahill M. Self-reported knowledge, attitudes and behaviour towards evidence-based practice of occupational therapy students in Ireland: evidence-based practice, Irish OT students. Occup Ther Int 2012;19(1):7–16.
76. Van Mullem C, Burke LJ, Dohmeyer K, Farrell M, Harvey S, John L, *et al*. Strategic planning for research use in nursing practice. JONA J Nurs Adm 1999;29(12):38–45.
77. Eller LS, Kleber E, Wang SL. Research knowledge, attitudes and practices of health professionals. Nurs Outlook 2003;51(4):165–70.
78. Upton D, Lewis B. Clinical effectiveness and EBP: design of a questionnaire. Br J Ther Rehabil 1998;5(12):647–50.
79. Upton D. Clinical effectiveness and EBP 2: attitudes of health-care professionals. Br J Ther Rehabil 1999;6(1):26–30.
80. Jette DU, Bacon K, Batty C, Carlson M, Ferland A, Hemingway RD, *et al*. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. Phys Ther 2003;83(9):786–805.
81. Hendricson WD, Rugh JD, Hatch JP, Stark DL, Deahl T, Wallmann ER. Validation of an instrument to assess evidence-based practice knowledge, attitudes, access, and confidence in the dental environment. J Dent Educ 2011;75(2):131–44.
82. Cobban S, Profetto-McGrath J. Knowledge sources used by Alberta dental hygienists: a pilot study. Can J Dent Hyg 2007;41(4):176.
83. Veeramah V. Utilization of research findings by graduate nurses and midwives. J Adv Nurs 2004;47(2):183–91.
84. Chernick L, Pusic M, Liu H, Vazquez H, Kwok M. A pediatrics-based instrument for assessing resident education in evidence-based practice. Acad Pediatr 2010;10(4):260–5.
85. Filippini A, Sessa A, Giuseppe GD, Angelillo IF. Evidence-based practice among nurses in Italy. Eval Health Prof 2011;34(3):371–82.
86. Kahveci R, Meads C. Is primary care evidence-based in Turkey? A cross-sectional survey of 375 primary care physicians. J Evid-Based Med 2009;2(4):242–51.
87. Lavis JN, Guindon GE, Cameron D, Boupha B, Dejman M, Osei EJ, *et al*. Bridging the gaps between research, policy and practice in low-and middle-income countries: a survey of researchers. CMAJ 2010;182(9):E350–61.
88. Bennett KJ, Sackett DL, Haynes RB, Neufeld VR, Tugwell P, Roberts R. A controlled trial of teaching critical appraisal of the clinical literature to medical students. JAMA 1987;257(18):2451–4.
89. Weberschock TB, Ginn TC, Reinhold J, Strametz R, Krug D, Bergold M, *et al*. Change in knowledge and skills of Year 3 undergraduates in evidence-based medicine seminars. Med Educ 2005;39(7):665–71.
90. Haynes RB, McKibbon KA, Walker CJ, Ryan N, Fitzgerald D, Ramsden MF. Online access to MEDLINE in clinical settings: a study of use and usefulness. Ann Intern Med 1990;112(1):78–84.
91. Haynes RB, Johnston ME, McKibbon KA, Walker CJ, Willan AR. A program to enhance clinical use of MEDLINE. A randomized controlled trial. Online J Curr Clin Trials 1993:4005.
92. McKibbon KA, Haynes RB, Dilks CJW, Ramsden MF, Ryan NC, Baker L, *et al*. How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Comput Biomed Res 1990;23(6):583–93.
93. Biondi-Zoccai G, editor. Umbrella reviews: evidence synthesis with overviews of reviews and meta-epidemiologic studies [internet]. Cham: Springer International Publishing; 2016 [cited 2021 Jan 12]. Available from: http://link.springer.com/10.1007/978-3-319-25655-9.
94. Terwee CB, Vet HCWde, Mokkink LB. Reply to `COSMIN for quality rating systematic reviews on psychometric properties'. Phys Ther Rev 2015;20(4):266–7.
95. Rethlefsen ML, Farrell AM, Trzasko LCO, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol 2015;68(6):617–26.
96. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, *et al*. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1):39.
97. COSMIN. A protocol guideline for systematic reviews of PROMs [Internet]. n.d. [cited 2021 Feb 1]. COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). Available from: http://www.cosmin.nl.
98. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ 2015;49(6):560–75.
99. St-Onge C, Young M, Eva K, Hodges B. Validity: one word with a plurality of meanings. Adv Health Sci Educ 2017;22(4):853–67.
100. Kuspinar A, Mayo NE. A Review of the psychometric properties of generic utility measures in multiple sclerosis. PharmacoEconomics 2014;32(8):759–73.

101. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, 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.
102. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What Is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10(s2):S94–105.
103. Khadjesari Z, Boufkhed S, Vitoratou S, Schatte L, Ziemann A, Daskalopoulou C, *et al*. Implementation outcome instruments for use in physical healthcare settings: a systematic review. Implement Sci 2020;15(1):66.
104. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? J Gen Intern Med 2005;20(12):1159–64.
105. Shaw S. Tracing the evolution of validity in educational measurement: past issues and contemporary challenges. Res Matters 2011;11:6.
106. Van de Mortel TF. Faking It: social desirability response bias in self-report research. Aust J Adv Nurs 2008;25 (4):40–8.
107. Raphael K. Recall Bias: A proposal for assessment and control. Int J Epidemiol 1987;16:167–70.
108. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006;119(2.):166.e7–166.e16.
109. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567(7748):305–7.
110. Saunders H, Gallagher-Ford L, Kvist T, Vehvilainen-Julkunen K. Practicing healthcare professionals' evidence-based practice competencies: an overview of systematic reviews. Worldviews Evid Based Nurs 2019;16(3):176–85.
111. Butcher NJ, Monsour A, Mokkink LB, Terwee CB, Tricco AC, Gagnier J, *et al.* Needed: guidance for reporting knowledge synthesis studies on measurement properties of outcome measurement instruments in health research. BMJ Open [internet]. 2021 [cited 2021 Feb 1]. Available from: https://bmjopen.bmj.com/content/needed-guidance-reporting-knowledge-synthesis-studies-measurement-properties-outcome.
112. Messick S. Validity. In: Educational measurement American Council on Education series on higher education. 3rd ed Phoenix, Arizona: Oryx Press; 1993.
113. Kane MT. Current concerns in validity theory. J Educ Meas 2001;38(4):319–42.
114. Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. Qual Life Res 2018;27(7):1695–710.
115. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? J Gen Intern Med 2004;19(9):971–7.
116. Terwee CB, Prinsen CA, Chiarotto A, De Vet H, Bouter LM, Alonso J, *et al*. COSMIN methodology for assessing the content validity of PROMs-user manual [internet]. 2018 [cited 2021 Feb 01]. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf.
117. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, de Vet HC, *et al*. User manual - COSMIN methodology for systematic reviews ofpatient-reported outcome measures (PROMs). User manual 2018;78:1.
118. McHugh M. Interrater reliability: the Kappa statistic. Biochem Med 2012;22(3):276–82.

Appendix I: Search strategy

MEDLINE (Ovid)

Search initially conducted on May 8, 2020. Search run again during analysis on January 18, 2021.

Search	Query	Records retrieved
#1	Students/ or exp Education, Medical/ or exp Education, Continuing/ or exp Competency-Based Education/ or exp Education, Professional/ or exp Education, Distance/ or exp Education, Professional, Retraining/ or exp Education, Medical, Continuing/ or education.fs. OR physical Therapy Modalities/ or "Physical Therapy (Specialty)"/ or Physical Therapists/ OR (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*).tw,kf. OR occupational therapy/ OR Nursing Education Research/ or exp Education, Nursing/ OR exp Physicians/ or (physician* or doctor*).tw,kf. OR exp Nurses/ or nurse*.tw,kf. OR (health care professional* or health professional* or healthcare professional*). tw,kf. OR Speech-Language Pathology/ OR (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*").tw,kf. OR Allied Health Occupations/ OR allied health.tw,kf. OR nursing.ti.	1,660,302
#2	exp Evidence-Based Practice/ OR (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making).tw,kf. OR (question formulation or critical appraisal).tw,kf. OR (research adj2 (update or "use")).tw,kf.	127,854
#3	"Surveys and Questionnaires"/ OR Psychometrics/ or Educational measurement/ OR (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*).tw,kf.	6,561,843
#4	Health Knowledge, Attitudes, Practice/ OR (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*).tw,kf.	5,898,290
#5	*CADTH systematic review fi/ter*meta-analysis.pt. OR meta-analysis/ or systematic review/ or meta-analysis as topic/ or "meta analysis (topic)"/or"systematic review (topic)"/ OR ((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf,kw. OR ((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,kf,kw. OR ((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kf,kw. OR (data synthes* or data extraction* or data abstraction*).ti,ab,kf,kw. OR handsearch* or hand search*).ti,ab,kf,kw. OR mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kf,kw. OR met analy* or metanaly* or technology assessment* or HTA or HTAs or technology overview* or technology appraisal*).ti, ab,kf,kw. OR (meta regression* or metaregression*).ti,ab,kf,kw. OR (meta-analy* or metaanaly* or systematic review* or biomedical technology assessment* or bio-medical technology assessment*).mp,hw. OR (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw. OR (cochrane or (health adj2 technology assessment) or evidence report).jw. OR (meta-analysis or systematic review).md. OR (comparative adj3 (efficacy or effectiveness)).ti,ab,kf,kw.OR (outcomes research or relative effectiveness).ti,ab,kf,kw. OR ((indirect or indirect treatment or mixed-treatment) adj comparison*).ti,ab,kf,kw.	501,979
#6	#1 AND #2 AND #3 AND #4 AND #5	1634
No language limitations in the search.		

Embase (Ovid)

Search initially conducted on May 8, 2020. Search ran again during analysis on January 18, 2021.

Search	Query	Records retrieved
#1	1. *student/ 2. exp *medical education/ 3. exp *continuing education/ 4. *physiotherapy/ 5. (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*).tw,kw. 6. *occupational therapy/ 7. exp *nursing education/ 8. exp *physician/ 9. (physician* or doctor*).tw,kw. 10. exp *nurse/ 11. nurse*.tw,kw. 12. (health care professional* or health professional* or healthcare professional*).tw,kw. 13. (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*").tw,kw. 14. allied health.tw,kw. 15. nursing.ti. 16.1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	1,505,482

(*Continued*)		
Search	**Query**	**Records retrieved**
#2	17. exp* evidence based practice/ 18. (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making).tw,kw. 19. (question formulation or critical appraisal).tw,kw. 20. (research adj2 (update or "use")).tw,kw. 21. 17 or 18 or 19 or 2	134,588
#3	22. *questionnaire/ 23. *psychometry/ 24. (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*).tw,kw. 25. 22 or 23 or 24	7,250,904
#4	26. (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*).tw,kw.	6,470,789
#5	CADTH systematic review fi/ter 28. meta-analysis.pt. 29. meta-analysis/ or systematic review/ or meta-analysis as topic/ or "meta analysis (topic)"/or "systematic review (topic)"/ 30. ((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kw. 31. ((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,kw. 32. ((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kw. 33. (data synthes* or data extraction* or data abstraction*).ti,ab,kw. 34. (handsearch* or hand search*).ti,ab,kw. 35. (mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kw. 36. (met analy* or metanaly* or technology assessment* or HTA or HTAs or technology overview* or technology appraisal*).ti,ab,kw. 37. (meta regression* or metaregression*).ti,ab,kw. 38. (meta-analy* or metaanaly* or systematic review* or biomedical technology assessment* or bio-medical technology assessment*).mp,hw. 39. (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw. 40. (cochrane or (health adj2 technology assessment) or evidence report).jw. 41. (meta-analysis or systematic.md. 42. (comparative adj3 (efficacy or effectiveness)).ti,ab,kw. 43. (outcomes research or relative effectiveness).ti,ab,kw. 44. ((indirect or indirect treatment or mixed-treatment) adj comparison*).ti,ab,kw.	721,283
#6	All categories combined by AND	1986
No language limitations in the search.		

PsycINFO (Ovid)

Search initially conducted on May 8, 2020. Search ran again during analysis on January 18, 2021.

Search	Query	Records retrieved
#1	1. exp Students/ 2. exp Medical Education/ 3. exp Continuing Education/ 4. exp Professional Competence/ 5. exp Physical Therapy/ 6. (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*).tw,id. 7. exp Occupational Therapy/ 8. exp Nursing Education/ 9. exp Physicians/ 10. (physician* or doctor*).tw,id. 11. exp Nurses/ 12. (health care professional* or health professional* or healthcare professional*).tw,id. 13. exp Speech Language Pathology/ 14. (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*").tw,id. 15. exp Allied Health Personnel/ 16. allied health.tw,id. 17. nursing.ti. 18. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	468,234

(*Continued*)

Search	Query	Records retrieved
#2	19. exp Evidence Based Practice/ 20. (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making).tw,id. 21. (question formulation or critical appraisal).tw,id. 22. (research adj2 (update or "use")).tw,id. 23. 19 or 20 or 21 or 22	29,616
#3	24. exp Surveys/ or Questionnaires/ 25. exp Psychometrics/ 26. exp Educational Measurement/ 27. (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*).tw,id. 28. 24 or 25 or 26 or 27	1,494,090
#4	29. exp Health Attitudes/ or exp Health Personnel Attitudes/ or exp Health Behavior/ or exp Knowledge Level/ or exp Health Knowledge/ 30. (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*).tw,id. 31. 29 or 30	1,873,096
#5	CADTH systematic review filter 33. meta-analysis.pt. 34. meta-analysis/ or systematic review/ or meta-analysis as topic/ or "meta analysis (topic)"/or "systematic review (topic)"/ 35. ((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,id. 36. ((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,id. 37. ((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,id. 38. (data synthes* or data extraction* or data abstraction*).ti,ab,id. 39. (handsearch* or hand search*).ti,ab,id. 40. (mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,id. 41. (met analy* or metanaly* or technology assessment* or HTA or HTAs or technology overview* or technology appraisal*).ti, ab,id. 42. (meta regression* or metaregression*).ti,ab,id. 43. (meta-analy* or metaanaly* or systematic review* or biomedical technology assessment* or bio-medical technology assessment*).mp,hw. 44. (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw. 45. (cochrane or (health adj2 technology assessment) or evidence report).jw. 46. (meta-analysis or systematic.md. 47. (comparative adj3 (efficacy or effectiveness)).ti,ab,id. 48. (outcomes research or relative effectiveness).ti,ab,id. 49. ((indirect or indirect treatment or mixed-treatment) adj comparison*).ti,ab,id. 50. or/33-49	102,346
#6	18 AND 23 AND 28 AND 31 AND 50	359
No language limitations in the search.		

CINAHL (EBSCO)

Search initially conducted on May 8, 2020. Search ran again during analysis on January 18, 2021.

Search	Query	Records retrieved
#1	(MH "Students") (MH "Education, Health Sciences") (MH "Education, Continuing") (MH "Education, Competency-Based") (MH "Physical Therapy+") (MH "Occupational Therapy+") AB (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*)or TI (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*) (MH "Education, Nursing+") (MH "Physicians+") AB (physician* or doctor*) (MH "Nurses+") AB (health care professional* or health professional* or healthcare professional*) or TI (health care professional* or health professional* or healthcare professional*) (MH "Speech-Language Pathology") AB (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*") or TI (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*") (MH "Allied Health Professions") AB "Allied Health" or TI "Allied Health" TI "nursing" or AB (nursing) *All of the above lines were combined with OR.*	1,098,347
#2	(MH "Professional Practice, Evidence-Based+") AB (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making) or TI (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making) AB (question formulation or critical appraisal) or TI (question formulation or critical appraisal) research N3 uptake research N3 use *All ofthe above lines were combined with OR.*	104,064
#3	(MH "Surveys")OR (MH "Structured Questionnaires")OR (MH "Open-Ended Questionnaires")OR (MH "Survey Research") (MH "Psychometrics") (MH "Educational Measurement") TI (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*) or AB (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*) *All ofthe above lines were combined with OR.*	1,501,759
#4	(MH "Attitude to Health")OR(MH"Attitude of Health Personnel") (MH "Professional Knowledge") (MH "Professional Practice, Research-Based")OR (MH "Professional Practice, Theory-Based")OR (MH "Practice Patterns") TI (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*) or AB (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*) *All ofthe above lines were combined with OR.*	1,365,173
#5	*CADTH systematic review filter* Meta analysis/ TI meta analysis or meta analyses OR AB meta analysis or meta analyses (MH "Literature Review+") TI systematic N1 (review or overview) OR AB systematic N1 (review or overview) OR (MH "Systematic Review") *All ofthe above lines were combined with OR.* *NOT* PT commentary PT letter PT editorial (MH "Animals")	177,322
#6	#1 AND #2 AND #3 AND #4 AND #5	1228
No language limitations in the search.		

ERIC (EBSCO)

Search initially conducted on May 8, 2020. Search ran again during analysis on January 18, 2021.

Search	Query	Records retrieved
#1	TI "nursing" or AB (nursing)AB "Allied Health" or TI "Allied Health"AB (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*") or TI (speech language pathologist* or speech language therapist* or speech pathologist* or speech therapist* or "speech and language pathologist*" or "speech and language therapist*")DE "Speech Language Pathology"AB (health care professional* or health professional* or healthcare professional*) or TI (health care professional* or health professional* or healthcare professional*)DE "Nurses"AB (physician* or doctor*) or TI (physician* or doctor*)DE "Physicians"DE "Nursing Education"AB (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*) or TI (medical student* or resident* or trainee* or rehabilitation science* or physical therap* or physiotherap* or occupational therap*)((((DE "Physical Therapy")OR(DE"Occupational Therapy")) OR (DE "Allied Health Occupations")) OR (DE "Nursing")) OR (DE "Medicine")DE "Continuing Education"(DE "Allied Health Occupations Education") OR(DE"Medical Education")DE "Students"*All of the above lines were combined with OR.*	86,545
#2	DE "Theory Practice Relationship"research N3 useresearch N3 uptakeAB (question formulation or critical appraisal) or TI (question formulation or critical appraisal)AB (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making) or TI (Evidence based medicine or evidence based practice or EBP or EBM or evidence informed decision making)DE "Evidence Based Practice"*All of the above lines were combined with OR.*	30,146
#3	TI (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*) or AB (Instrument* or survey* or questionnaire* or scale* or tool* or measure* or evaluation* or validity or reliability or psychometric*)DE "Educational Assessment"DE "Psychometrics"(DE "Surveys")OR(DE"Questionnaires")*All of the above lines were combined with OR.*	589,087
#4	TI (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*) or AB (skill* or behavior* or behaviour* or knowledge* or attitude* or competenc* or attribute* or factor* or resource*)((DE "Knowledge Level") OR (DE "Attitudes")) OR (DE "Behavior")*All of the above lines were combined with OR.*	728,313
#6	#1 AND #2 AND #3 AND #4	485
No language limitations in the search. SR filter was judged to be unnecessary for ERIC.		

Appendix II: Data extraction form

Characteristics of reviews	
Title of article	
Authors	
Publication year	
Country of origin (primary author)	
Population (eg, residents, physiotherapists)	
EBP context (clinical practice, education, or both)	
Setting (eg, rehabilitation center)	
Aim of the systematic review	
Number of included primary studies	
Number of included measures	
Number of identifiable measures	
Names of all identifiable measures, with primary study reference	
Critical appraisal instrument used to assess primary studies within the review	
Quality assessment of primary studies (results)	
Did the authors assess adequacy of included measures?	
Instrument used to assess adequacy of included measures	
Descriptive info on measures reported in systematic review (frequencies, etc.)	
Conceptualization of EBP (and reference)	
Reported outcomes	
Categorization of measures	
Reported strengths of systematic review	
Reported limitations of the systematic review	

Search methods	
Databases searched	
Search time frame	
Search language limits	
EBP domains included	
# of independent reviewers for study eligibility at title/abstract and full text	
Was detailed syntax for one database at least available in the publication? (yes/no)	
Were the searches conducted by a library/information specialist? (yes/no)	
Was reference checking performed? (yes/no)	
Were reasons for excluding articles reported? (yes/no)	
Inclusion: Study design	
Inclusion: Population	
Inclusion: Outcomes	
Inclusion: Type of measure	
Inclusion: EBP domain	
Exclusion criteria (if reported)	

Additional information on the methods used	
Protocol registered a priori (yes/no)?	
Reference to a systematic review methodology (yes/no) and which one?	

(*Continued*)	
Additional information on the methods used	
Additional information on the methods used	
Reference to a reporting guideline (yes/no) and which one?	
PRISMA flowchart (or equivalent) (yes/no) and which one?	
Included PRISMA checklist (or equivalent) (yes/no) and which one?	

Reporting	
Psychometric properties and characteristics	
Describe the depth and breadth of reporting of the below psychometric properties* and characteristics. The depth and breadth are defined as the amount of detail that the authors provide (eg, checklist, "established," frequency, keyword, information on study methods, population, raw data/scores, conclusions). Also, precisely whether the information is presented for all or a subset of measures and whether the detail provided is variable across properties and measures. Mention whether this information is provided in a supplementary file. Also, did any reviews present pooled estimates of measurement properties or significance tests?	
• Content validity (including information on the measure development process and on the theoretical basis or conceptual framework) • Structural validity (eg, confirmatory Factor Analysis or Tucker Lewis index; factor analysis or IRT/ Rasch analysis) • Internal consistency (eg, Cronbach's alpha) • Cross-cultural validity (ie, degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM) • Reliability (eg, ICC or weighted kappa) • Measurement error (ie, the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured; eg, limits of agreement, smallest detectable change, minimally important change) • Criterion validity (ie, degree to which the scores of a PROM are an adequate reflection of a "gold standard") • Hypothesis testing for construct validity (eg, the comparison instrument(s) that is/are used to compare the PROM(s) against, and the expected direction and magnitude of the correlation) • Responsiveness (ie, ability of a PROM to detect change over time in the construct to be measured) • Interpretability (ie, degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to a PROM's quantitative scores or change in scores [eg, missing total scores, missing items, floor and ceiling effects, distribution of scores]) • Feasibility (eg, time required to administer measure, time required to score, missing data, expertise required to score, cost to administer and score, administrative support required, comprehensibility)	
Are there any properties or characteristics of measures that do not align with the above categories?	
Is the reporting of psychometric evidence categorized by primary study or by measure?	

Adequacy assessment results reported	
At the property level? For all measures or a subset?	
At the measure level? For all measures or a subset?	

Method of assessing the adequacy of measures	
Tool/standard referenced	
Measurement properties (and scoring) associated with the tool/standard**	
Measurement properties used in the systematic review	
Number of assessors	

(*Continued*)	
Method of assessing the adequacy of measures	
Criteria of adequacy at property level	
Criteria of adequacy at measure level	

Recommendations concerning EBP measures	
Names of recommended measures and ranking (if applicable)	
Recommendations concerning EBP measures	
Names of measures deemed adequate or high quality and ranking (if applicable)	
Rationale for recommendations (if provided)	
EBP domain(s) and/or EBP step associated with recommended, high-quality, or adequate measures (as reported in the systematic review)	
Population in which the measure was developed (health care profession, level, number of participants, country) associated with primary study	
Scale construction (units, scoring range)	
Format of measure (ie, written, web based, self-report survey, audit)	

EBP, evidence-based practice; ICC, intraclass correlation coefficient; IRT, item response theory; PROM, patient-reported outcome measure.
*Information on the reporting of measure characteristics was extracted by mapping the depth and breadth of reporting on to the 11 properties taken from the widely recognized COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Guideline for Systematic Reviews of Outcome Measurement Instruments.[117] Although interpretability and feasibility are not formally measurement properties, they are crucial elements in determining the suitability of a measure.
**This information was extracted by accessing the references from citations in the systematic reviews.

Appendix III: Studies ineligible following full-text review

	Reference	Reason for exclusion
1	Adam K, Peters S, Chipchase L. Knowledge, skills and professional behaviours required by occupational therapist and physiotherapist beginning practitioners in work-related practice: a systematic review. Aust Occup Ther J. 2013;60(2):76-84.	No specific instruments reported
2	Aglen B. Pedagogical strategies to teach bachelor students evidence-based practice: a systematic review. Nurse Educ Today. 2016;36:255-63.	No specific instruments reported
3	Ahmadi S-F, Baradaran HR, Ahmadi E. Effectiveness of teaching evidence-based medicine to undergraduate medical students: a BEME systematic review. Med Teach. 2015;37(1):21-30.	No specific instruments reported
4	Albarqouni L, Hoffmann T, Straus S, Olsen NR, Young T, Ilic D, *et al.* Core competencies in evidence-based practice for health professionals: consensus statement based on a systematic review and Delphi survey. JAMA Netw Open. 2018;1(2):e180281.	Not a systematic review
5	Athanasakis E. Nurses' research behavior and barriers to research utilization into clinical nursing practice: a closer look. ResearchGate. 2013.	Not a systematic review
6	Baatiema L, Otim ME, Mnatzaganian G, de-Graft Aikins A, Coombes J, Somerset S. Health professionals' views on the barriers and enablers to evidence-based practice for acute stroke care: a systematic review. Implement Sci. 2017;12(1):74/.	No specific instruments reported
7	Barrett J, Gonsalvez CJ, Shires A. Evidence-based practice within supervision during psychology practitioner training: a systematic review. Clin Psychol. 2019;24(1):3-17.	No specific instruments reported
8	Barzkar F, Baradaran HR, Koohpayehzadeh Jalil. Knowledge, attitudes and practice of physicians toward evidence-based medicine: a systematic review. J Evid Based Med. 2018;11(4):246-51.	No specific instruments reported
9	Brettle A. Information skills training: a systematic review of the literature. Health Info Lib J. 2003;20(Suppl 1):3-9.	No specific instruments reported
10	Busari JO, Stammen LA, Gennissen LM, Moonen RM. Evaluating medical residents as managers of care: a critical appraisal of assessment methods. Adv Med Educ Pract. 2014;5:27-37.	Not about EBP
11	Camargo FC, Iwamoto HH, Galvão CM, de Araujo Pereira G, Andrade RB, Masso GC. Competences and barriers for the evidence-based practice in nursing: an integrative review. Rev Bras Enferm. 2018;71(4):2030-8.	No specific instruments reported
12	Camargo FC, Iwamoto HH, Galvão CM, de Arauijo Pereira G, Andrade RB, Masso GC. Competences and barriers for the Evidence-Based Practice in Nursing: an integrative review. Rev Bras Enferm. 2018;71(4):2030-8.	Duplicate
13	Campbell A, Louie-Poon S, Slater L, Scott SD. Knowledge translation strategies used by healthcare professionals in child health settings: an updated systematic review. J Pediatr Nurs. 2019;47:114-20.	No specific instruments reported
14	Cartabellotta A, Montalto G, Notarbartolo A. Levels of scientific evidence and strength of clinical recommendations. From trials to guidelines. The Italian Group on Medicine Based on Evidence-(GIMBE). Recenti Prog Med. 1997;88(7-8):342-7.	Not a systematic review
15	Carter AG, Creedy DK, Sidebotham M. Efficacy of teaching methods used to develop critical thinking in nursing and midwifery undergraduate students: a systematic review of the literature. Nurse Educ Today. 2016;40:209-18.	Not about EBP
16	Chan A, Purcell A, Power E. A systematic review of assessment and intervention strategies for effective clinical communication in culturally and linguistically diverse students. Med Educ. 2016;50(9):898-911.	Not about EBP
17	Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. BMJ. 2004;329(7473):1017.	No specific instruments reported
18	Cui C, Li Y, Geng D, Zhang H, Jin C. The effectiveness of evidence-based nursing on development of nursing students' critical thinking: a meta-analysis. Nurse Educ Today. 2018;65:46-53.	Not about EBP
19	da Silva TM, Costa LdCM, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. Man Ther. 2015;20(3):388-401.	No specific instruments reported
20	de Wit K, Curran J, Thoma B, Dowling S, Lang E, Kuljic N, *et al.* Review of implementation strategies to change healthcare provider behaviour in the emergency department. CJEM. 2018;20(3):453-60.	No specific instruments reported
21	Dizon J, Grimmer-Somers KA, Kumar S. Current evidence on evidence-based practice training in allied health: a systematic review of the literature. Int J Evid Based Healthc. 2012;10(4):347-60.	No specific instruments reported
22	Escoffery C; Lebow-Skelley E, Haardoerfer R, Boing E, Udelson H, Wood R, *et al.* A systematic review of adaptations of evidence-based public health interventions globally. Implement Sci. 2018;13(1):125.	Not about EBP
23	Estabrooks CA, Floyd JA, Scott-Findlay S, O'Leary KA, Gushta M. Individual determinants of research utilization: a systematic review. 2003;43(5):506-20.	Measures of research utilization
24	Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a systematic review. BMC Health Serv Res. 2007;7:119.	No specific instruments reported

(*Continued*)		
	Reference	**Reason for exclusion**
25	Forrest JL, Miller SA. Evidence-based decision making in dental hygiene education, practice, and research. J Dent Hyg. 2001;75(1):50-63.	Not a systematic review
26	Gao W, Jin Y, Sun M. Meta analysis of influence of evidence-based nursing teaching on learning effect of nursing students. 2010. [full reference not available]	No full text available
27	Garg A, Turtle KM. Effectiveness of training health professionals in literature search skills using electronic health databases-a critical appraisal. Health Info Lib J. 2003;20(1):33-41.	No specific instruments reported
28	Geerts JM, Goodall AH, Agius S. Evidence-based leadership development for physicians: a systematic literature review. Soc Sci Med. 2020;246:112709.	Not about EBP
29	Ginossar T, Heckman CJ, Cragun D, Quintiliani LM, Proctor EK, Chambers DA, *et al.* Bridging the chasm: challenges, opportunities, and resources for integrating a dissemination and implementation science curriculum into medical education. J Med Educ Curric Dev. 2018;5:2382120518761875.	No specific instruments reported
30	Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. Acad Med. 1999;74(6):686-94.	Not a systematic review
31	Greenhalgh T. Integrating qualitative research into evidence based practice. Endocrinol Metab Clin North Am. 2002; 31(3):583-601.	Not a systematic review
32	Haggman-Laitila A. Evidence based nursing: systematic review of implementation. ResearchGate. 2009.	No full text available
33	Haggman-Laitila A. Factors facilitating evidence based nursing - systematic review of nurses' perceptions. 2009. [full reference not available]	No full text available
34	Haggman-Laitila A. Promoting evidence based nursing - systematic review of conceptual models. 2009. [full reference not available]	No full text available
35	Harris J, Kearley K, Heneghan C, Meats E, Roberts N, Perera R, *et al.* Are journal clubs effective in supporting evidence-based decision making? A systematic review. BEME Guide No. 16. Med Teach. 2011;33(1):9-23.	No specific instruments reported
36	Hart MD. Informatics competency and development within the US nursing population workforce: a systematic literature review. Comput Inform Nurs. 2008;26(6):320-9.	Not about EBP
37	Hebert RS, Levine RB, Smith CG, Wright SM. A systematic review of resident research curricula. Acad Med. 2003; 78(1):61-8.	No specific instruments reported
38	Hecht L, Buhse S, Meyer G. Effectiveness of training in evidence-based medicine skills for healthcare professionals: a systematic review. BMC Med Educ. 2016;16:103.	No specific instruments reported
39	Hines S, Ramsbotham J, Coyer F. The effectiveness of interventions for improving the research literacy of nurses: a systematic review. Worlviews Evid Based Nurs. 2015;12(5):265-72.	No specific instruments reported
40	Hines S, Ramsbotham J, Coyer F. Interventions for improving the research literacy of nurses: a systematic review. JBI Database System Rev Implement Rep. 2016;14(2):256-94.	No specific instruments reported
41	Hoegen P, De Bot C, Echteld M, Vermeulen H. 69 EBP related self-efficacy among healthcare and social care professionals: a systematic review. BMJ Evid Based Med. 2018;23:A34.	No full text available
42	Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart R. Teaching critical appraisal skills in healthcare settings. Cochrane Database Syst Rev. 2011;(11):CD001270.	No specific instruments reported
43	Horsley T, O'Neill J, McGowan J, Perrier L, Kane G, Campbell C. Interventions to improve question formulation in professional practice and self-directed learning. Cochrane Database Syst Rev. 2010;(5):CD00735.	No specific instruments reported
44	Ilic D, de Voogt A, Oldroyd J. The use of journal clubs to teach evidence-based medicine to health professionals: a systematic review and meta-analysis. J Evid Based Med. 2020;13(1):42-56.	No specific instruments reported
45	Ilic D, Maloney S. Methods of teaching medical trainees evidence-based medicine: a systematic review. Med Educ. 2014;48(2):124-35.	No specific instruments reported
46	Kahwati L, Carmody D, Berkman N, Sullivan HW, Aikin KJ, DeFrank J. Prescribers' knowledge and skills for interpreting research results: a systematic review. J Contin Educ Health Prof. 2017;37(2):129-36.	No specific instruments reported
47	Kajermo KN, Bostrom A-M, Thompson DS, Hutchinson AM, Estabrooks CA, Wallin L. The BARRIERS scale - the barriers to research utilization scale: a systematic review. Implement Sci. 2010;5:32.	Measures of research utilization
48	Khalifah AM, Celenza A. Teaching and assessment of dentist-patient communication skills: a systematic review to identify best-evidence methods. J Dent Educ. 2019;83(1):16-31.	Not about EBP
49	Komprood SR. Nursing student attitudes toward oncology nursing: an evidence-based literature review. Clin J Oncol Nurs. 2013;17(1):E21-8.	Not a systematic review

(*Continued*)		
	Reference	**Reason for exclusion**
50	Kyriakoulis K, Patelarou A, Laliotis A, Wan AC, Matalliotakis M, Tsiou C, *et al.* Educational strategies for teaching evidence-based practice to undergraduate health students: systematic review. J Educ Eval Health Prof. 2016;13-34.	No specific instruments reported
51	Legare F, Moher D, Elwyn G, LeBlanc A, Gravel K. Instruments to assess the perception of physicians in the decision-making process of specific clinical encounters: a systematic review. BMC Med Inform Decis Mak. 2007;7:30.	Not about EBP
52	Lizarondo L, Grimmer-Somers K, Kumar S. A systematic review of the individual determinants of research evidence use in allied health. J Multidiscip Healthc. 2011;4:261-72.	No specific instruments reported
53	Maggio LA, Kung JY. How are medical students trained to locate biomedical information to practice evidence-based medicine? A review of the 2007-2012 literature. J Med Libr Assoc. 2014;102(3):184-91.	No specific instruments reported
54	Mazmanian PE, Davis DA, Galbraith R, American College of Chest Physicians, Health; Science Policy, Committee. Continuing medical education effect on clinical outcomes: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. Chest. 2009;135(Suppl 3):49S-55S.	Not about EBP
55	Meijers JMM, Janssen MAP, Cummings GG, Wallin L, Estabrooks CA, Halfens RYG. Assessing the relationships between contextual factors and research utilization in nursing: systematic literature review. J Adv Nurs. 2006;55 (5):622-35.	No specific instruments reported
56	Menon A, Korner-Bitensky N, Kastner M, McKibbon KA, Straus S. Strategies for rehabilitation professionals to move evidence-based knowledge into practice: a systematic review. J Rehabil Med. 2009;41(13):1024-32.	No specific instruments reported
57	Myers CT, Lotz J. Practitioner training for use of evidence-based practice in occupational therapy. Occup Ther Health Care. 2017;31(3):214-37.	No specific instruments reported
58	Newhouse RP. Evidence-based behavioral practice: an exemplar of interprofessional collaboration. J Nurs Adm. 2008;38(10):414-16.	Not a systematic review
59	Newhouse RP, White KM. Guiding implementation: frameworks and resources for evidence translation. J Nurs Adm. 2011;41(12):513-16.	Not a systematic review
60	Patelarou AE, Kyriakoulis KG, Stamou AA, Laliotis A, Sifaki-Pistolla D, Matalliotakis M. Approaches to teach evidence-based practice among health professionals: an overview of the existing evidence. Adv Med Educ Pract. 2017;8:455-64.	No specific instruments reported
61	Patelarou AE, Patelarou E, Brokalaki H, Dafermos V, Thiel L, Melas CD, *et al.* Current evidence on the attitudes, knowledge and perceptions of nurses regarding evidence-based practice implementation in European community settings: a systematic review. J Community Health Nurs. 2013;30(4):230-44.	No specific instruments reported
62	Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Hammick M, *et al.* A systematic review of how studies describe educational interventions for evidence-based practice: stage 1 of the development of a reporting guideline. BMC Med Educ. 2014;14:152.	No specific instruments reported
63	Ramis M-A, Chang A, Conway A, Lim D, Munday J, Nissen L. Theory-based strategies for teaching evidence-based practice to undergraduate health students: a systematic review. MBC Med Educ. 2019;19(1):267.	No specific instruments reported
64	Ramis MA, Chang A, Nissen L. Strategies for teaching evidence-based practice to undergraduate health students: a systematic review protocol. JBI Database System Rev Implement Rep. 2015;13(2):12-25.	No specific instruments reported
65	Ramis M-A, Chang A, Nissen L. Undergraduate health students' intention to use evidence-based practice after graduation: a systematic review of predictive modeling studies. Worldviews Evid Based Nurs. 2018;15(2):140-8.	No specific instruments reported
66	Ribeiro JP, Porto AR, Thofehrn MB. [Evidence-based practice: Methodological trends in nursing] Evidentia. 2012;9 (40). Spanish	Not English or French
67	Saunders H, Gallagher-Ford L, Kvist T, Vehvilainen-Julkunen K. Practicing healthcare professionals' evidence-based practice competencies: an overview of systematic reviews. Worldviews Evid Based Nurs. 2019;16(3):176-85.	No specific instruments reported
68	Simons MR, Zurynski Y, Cullis J, Morgan MK, Davidson AS. Does evidence-based medicine training improve doctors' knowledge, practice and patient outcomes? A systematic review of the evidence. Med Teach. 2019;41(5):532-8.	No specific instruments reported
69	Swanberg SM, Dennison CC, Farrell A, Machel V, Marton C, O'Brien KK, *et al.* Instructional methods used by health sciences librarians to teach evidence-based practice (EBP): a systematic review. J Med Libr Assoc. 2016;104(3):197-208.	No specific instruments reported
70	Thepwongsa I, Kirby C, Schattner P, Shaw J, Piterman L. Type 2 diabetes continuing medical education for general practitioners: what works? A systematic review. Diabet Med. 2014;31(12):1488-97.	Not about EBP
71	Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. BMJ Open. 2013;3(1):e001881.	No specific instruments reported

(*Continued*)		
	Reference	**Reason for exclusion**
72	Upton D, Stephens D, Williams B, Scurlock-Evans L. Occupational therapists' attitudes, knowledge, and implementation of evidence-based practice: a systematic review of published research. Br J Occup Res. 2014;77(1):24-38.	No specific instruments reported
73	Wallace J, Byrne C, Clarke M. Making evidence more wanted: a systematic review of facilitators to enhance the uptake of evidence from systematic reviews and meta-analyses. Int J Evid Based Healthc. 2012;10(4):338-46.	No specific instruments reported
74	Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. BMJ Open. 2012;2(5):e001220.	No specific instruments reported
75	Weerasekera P, Manring J, Lynn DJ. Psychotherapy training for residents: reconciling requirements with evidence-based, competency-focused practice. Acas Psychiatry. 2010;34(1):5-12.	Not a systematic review
76	Werb SB, Matear DW. Implementing evidence-based practice in undergraduate teaching clinics: a systematic review and recommendations. J Dent Educ. 2004;68(9):996-1003.	No specific instruments reported
77	Wieringa S, Greenhalgh T. 10 years of mindlines: a systematic review and commentary. Implement Sci. 2015;10:45.	No specific instruments reported
78	Wu Y, Brettle A, Zhou C, Ou J, Wang Y, Wang S. Do educational interventions aimed at nurses to support the implementation of evidence-based practice improve patient outcomes? A systematic review. Nurse Educ Today. 2018;70-109-14.	No specific instruments reported

EBP, evidence-based practice.

Appendix IV: Agreement between reviewers

A. Agreement for pilot screening at title and abstract level (n = 179)			
	Reviewer B		
Reviewer A	Include	Exclude	Row marginals
Include	5	12	17
Exclude	4	158	162
Column marginals	9	170	179

Proportionate agreement	Yes probability	No probability	Random agreement probability	Cohen's Kappa
0.91061	0.00478	0.85952	0.8643	0.34131

B. Agreement for *full set* screening at title and abstract level (n = 3393)			
	Reviewer B		
Reviewer A	Include	Exclude	Row marginals
Include	36	27	63
Exclude	7	3323	3330
Column marginals	43	3350	3393

Proportionate agreement	Yes probability	No probability	Random agreement probability	Cohen's Kappa
0.98998	0.00024	0.96899	0.96923	0.67434

C. Agreement for *full text* screening (n = 87)			
	Reviewer B		
Reviewer A	Include	Exclude	Row marginals
Include	10	5	15
Exclude	7	65	72
Column marginals	17	70	87

Proportionate Agreement	Yes Probability	No Probability	Random agreement probability	Cohen's Kappa
0.86207	0.03369	0.66587	0.69956	0.54089

Kappa statistics were calculated and interpreted according to the guidelines proposed by Cohen in his original paper on kappa.[30]

Interpretation of Kappa

Value of K	Strength of agreement
<0	Poor
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Based on Altman,[28] 1 999, and Landis & Koch,[29] 1977

Value of K	Level of agreement	% of data that are reliable
0-0.20	None	0-4%
0.21-0.39	Minimal	4-15%
0.40-0.59	Weak	15-35%
0.60-0.79	Moderate	35-63%
0.80-0.90	Strong	64-81%
Above 0.90	Almost perfect	82-100%

Based on McHugh,[118] 2012

Appendix V: All included evidence-based practice measures across the 10 systematic reviews

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Population	**All HCPs**	**Nurses**	**OTs**	**PTs**	**OTs**	**Medical students**	**Nurses**	**All HCPs**	**All HCPs**	**Fam med residents**		
Evidence-Based Practice Questionnaire (EBPQ) aka "revised Upton," "modified Upton and Lewis" (Upton & Upton,[61] 2006)		I	A		I		I	A (translation by Romani & Forni[63])			2 A, 3 I	5
Fresno (Ramos,[48] 2003)	A					A			A	I	3 A, 1 I	4
Taylor's questionnaire (Taylor,[53] 2001)	A			I		A			A		3 A, 1 I	4
Questionnaire on EBP and clinical effectiveness (Upton and Lewis,[78] 1998)			A	I	I		I with another version of the measure by authors				1 A, 3 I	4
Berlin Questionnaire (Fritsche *et al.*,[49] 2002)	A					A			A		3 A	3
MacRae *et al.*[51] (2004)	A					A			A		3 A	3
Philibert *et al.*[59] (2003)			A		I			A			2 A, 1 I	3
Barriers and Attitudes to Research in the Therapies "BART" (Metcalfe *et al.*[72] 2001)			A	I	I						1 A, 2 I	3

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Evidence-Based Practice Confidence Scale (EPIC) (Salbach and Jaglal, 2011)		I	I	I							3 I	3
Caldwell *et al.* (2007)			I		I		I				3 I	3
Pollock *et al.* (2000)			I	I	I						3 I	3
Melnyk *et al.* (2004)		I					I	I			3 I	3
Karlsson and Törnquist (2007)			I		I			I			3 I	3
ACE tool (Ilic *et al.*[8] 2014)	A					A					2 A	2
The Utrecht questionnaire (U-CEP) (Kortekaas *et al.*[60] 2017)	A					A					2 A	2
Evidence Based Practice Profile Questionnaire "EBPP" (McEvoy *et al.*[66] 2010)				I				A			1 A, 1 I	2
Edmonton Research Orientation Survey "EROS" (Pain *et al.*[73] 1996)			A		I						1 A, 1 I	2
Jette[80] (2003)				I				A			1 A, 1 I	2
Investigator-developed (Boström *et al.* 2009)							I	I			2 I	2
Adapted Fresno Test		I			I						2 I	2

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Developing Evidence-based Practice Questionnaire DEBPQ (Gerrish *et al.*, 2007)		I					I				2 I	2
Bennett *et al.*, 2003 (modified from McColl *et al.*, 1998)			I		I						2 I	2
Humphris *et al.* (2000) - Questionnaire based on a qualitative study and Pettingill *et al.* (1994)			I		I						2 I	2
Curtin and Jarama-zovic (2001) Questionnaire			I		I						2 I	2
Dysart and Tomlin (2002) Questionnaire			I		I						2 I	2
Pomeroy *et al.* (2003) Questionnaire			I		I						2 I	2
Sails *et al.* (2009) questionnaire developed based on Jette *et al.*[80] (2003) and Dysart and Tomlin (2002)			I		I						2 I	2
Tudiver OSCE (Tudiver *et al.* 2009)						I				I	2 I	2
Ross and Verdieck, (2003)									I	I	2 I	2
Fung *et al.* (2000)								I	I		2 I	2

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Straus *et al.* (2005); Lucas *et al.* (2004); Ellis *et al.* (1995)								I	I		2 I	2
Self-developed measure by Gerrish and Clayton (2004) (aka RUQ of Baessler)		I					I				2 I	2
Evidence-Based Practice Readiness Survey (EBPRS) (Thiel and Ghosh 2008)							I	I			2 I	2
Self-developed measure by Gerrish *et al.* (2011)		I					I				2 I	2
Barriers to Research Utilization Scale "BARRIERS" (Funk *et al.*[67] 1991)			A								1 A	1
Modified Knowledge, Attitude and Behaviour (KAB) questionnaire (Stronge and Cahill,[75] 2012)			A								1 A	1
The Knowledge, Attitudes, and Practices (KAP) Survey (Van Mullem *et al.*[76] 1999)			A								1 A	1
Chiu[65] (2009)								A			1 A	1

(Continued)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Evidence Based Practice Knowledge, Attitudes, Access and Confidence Evaluation (KACE) Hendricson *et al.*[81] (2011)								A			1 A	I
Cobban *et al.*[82] (2007)								A			1 A	1
Veeramah *et al.*[83] (2004)								A			1 A	1
Chernick *et al.*[84] (2010)								A			1 A	1
Fillipini *et al.*[85] (2011)								A			1 A	1
Kahveci and Meads[86] (2009)								A			1 A	1
Lavis *et al.* (2010)[87]								A			1 A	1
Bennett *et al.*[88] (1987)									A		1 A	1
Weberschock *et al.*[89] (2005)									A		1 A	1
Haynes *et al.*[91,90] (1990, 1993); McKibbon *et al.*[92] (1990)									A		1 A	1
The School Nursing EBP questionnaire (Adams, 2007)		I									1 I	1
Evidence-Based Nursing Attitude Questionnaire (EBNAQ) (Ruzafa-Martinez *et al.* 2011)		I									1 I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Yip *et al.* (2011)		I									1I	1
Quick EBP VIK (Values, Implementation, Knowledge) Survey		I									1 I	1
EBP measure developed by Majid *et al.* (2011)		I									1 I	1
Knowledge and Skills in Evidence-Based Nursing (KS-EBN)		I									1 I	1
Perceived EBP Knowledge Measure		I									1 I	1
Nurses' Attitudes Toward EBP Scale (NATES)		I									1 I	1
Modified Stevens EBP Readiness Inventory (ERI) (Finnish ERI)		I									1I	1
Johns Hopkins Nursing EBP Assessment Survey		I									1I	1
Persian translated EBP measure by Seyyedrasooli *et al.* (2012)		I									1I	1
Self-developed measure by Bostrom *et al.* (2013)		I									1I	1
Investigator-developed (based on Bostrom *et al.* 2009) by Florin *et al.* (2012)		I									1 I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
EBP Competency Tool (Melnyk *et al.*, 2018)		I									1 I	1
EBP-Implementation Scale (EBPI)		I									1I	1
EBP-Beliefs Scale (EBPB)		I									1 I	1
Modified Evidence-Based Nursing Education Questionnaire (EBEQ)		I									1I	1
Self-developed measure by Barako *et al.* (2012)		I									1I	1
Self-developed measure by Kim *et al.* (2013)		I									1 I	1
Modified Korean EBM questionnaire		I									1 I	1
Evidence-Based Practice Attitudes Scale (EBPAS)		I									1I	1
Attitudes to Evidence-Based Practice Questionnaire		I									1I	1
Single item measure for EBP knowledge		I									1I	1
Evidence-Based Practice Knowledge Assessment in Nursing (EKAN)		I									1I	1
Knowledge Assessment Test (KAT)		I									1I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Self-developed measure by Chiu *et al.*[64] (2010)		I									1I	1
Self-developed measure by Chew *et al.* (2015)		I									1I	1
Core Knowledge Questionnaire		I									1I	1
Graham *et al.* (2013)			I								1I	1
Bennett *et al.* (2007)			I								1I	1
Cameron *et al.* (2005)			I								1I	1
Cooke *et al.* (2008)			I								1I	1
Dopp *et al.* (2012)			I								1I	1
Gosling and Westbrook (2004)			I								1I	1
Heiwe *et al.* (2011) Questionnaire translated from Jette *et al.*[80] (2003) with some modifications			I								1I	1
Hu *et al.* (2012)			I								1I	1
Pain *et al.* (2004) General Use of Research			I								1I	1
Pain *et al.* (2004) Knowledge Acquisition Survey			I								1I	1
Pain *et al.* (2004) Individual semi-structured interviews			I								1I	1

(*Continued*)												
Summary of 204 unique measures identified across all included systematic reviews	**Albarqouni *et al.*[40] (2018)***	**Belita *et al.*[36] (2020)**	**Buchanan *et al.*[31] (2016)**	**Fernandez-Dommguez *et al.*[38] (2014)**	**Glegg and Holsti[32] (2010)**	**Kumaravel *et al.*[39] (2020)**	**Leung *et al.*[33] (2014)**	**Oude Rengerink *et al.*[37] (2013)***	**Shaneyfelt *et al.*[34] (2006)***	**Thomas and Kreptul[35] (2015)**	**Measures deemed adequate (A) by the authors versus those simply identified (I)**	**# of systematic reviews having identified the measure**
Powell and Case-Smith (2003) Questionnaire			I								1I	1
Gilman (2011)			I								1I	1
Sweetland and Craik (2001)			I								1I	1
McKenna *et al.* (2005)			I								1I	1
McCluskey (2003)			I								1I	1
Ramírez-Vélez *et al.* (2013)				I							1I	1
Bernhardsson and Larsson (2013)				I							1I	1
Palfreyman *et al.* (2003)				I							1I	1
Kamwendo (2002)				I							1I	1
Nilsagard and Lohse (2010)				I							1I	1
Fruth *et al.* (2010)				I							1I	1
Anthamatten (2009)				I							1I	1
Sabus (2008)				I							1I	1
Hadley *et al.* (2008)				I							1I	1
Bridges *et al.* (2007)				I							1I	1
Iles and Davidson (2006)				I							1I	1
van der Wees (2013)				I							1I	1
Cimoli (2012)				I							1I	1
Groth (2011)				I							1I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Salbach *et al.* (2007)				I							1I	1
Evidence-based Professional Practice Scale (EBPP-S) questionnaire				I							1I	1
Questions on EBP attitudes' questionnaire				I							1I	1
EBP Survey (self-report questionnaire by McCluskey and Lovarini 2005)					I						1I	1
BACES - Biostatistics and Clinical Epidemiology Skills assessment for medical residents by Barlow *et al.* (2015)						I					1I	1
Evidence-Based Medicine (EBM) test (Feldstein *et al.* 2010)						I					I	1
Educational pre-scription (Feldstein *et al.* 2009)						I					I	1
Mendiola-MCQ (Sanchez-Mendiola *et al.* 2012)						I					I	1
Frohna's OSCE (Frohna *et al.* 2006)						I					I	1
Foo *et al.* (2011)							I				1 I	1

(Continued)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Evidence Based Nursing Skills Assessment Tool (EBNSAT) (adapted from Lash *et al.* 2003) from Munroe *et al.* (2008)							I				1I	1
Evidence-Based Nursing Questionnaire (EBNQ) (Nagy *et al.* 2001)							I				1I	1
Yip Wai *et al.* (2013)							I				1 I	1
Mokhtar *et al.* (2012)							I				1 I	1
Filippini *et al.*[85] (2011)							I				1 I	1
Larrabee *et al.* (2007)							I				1 I	1
Egerod and Hansen (2005)							I				1 I	1
Research Awareness Questionnaire (RAQ) (McSherry *et al.* 2006)											1I	1
Carter and Stoecker (2006)								I			1 I	1
Ross (2010)								I			1 I	1
Shirkhedkar (2008)								I			1 I	1
Phua and Lim (2008)								I			1 I	1
Carney *et al.* (2004)								I			1 I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Forsman *et al.* (2010)								I			1 I	1
Olade (2004)								I			1 I	1
Lacey (1994)								I			1 I	1
Profetto-McGrath *et al.* (2010)								I			1 I	1
Abeysana *et al.* (2010)								I			1 I	1
Adams (2009)								I			1 I	1
Ahmad *et al.* (2009)								I			1 I	1
Tranmer *et al.* (2002)								I			1 I	1
Johansson *et al.* (2010)								I			1 I	1
Wallin (2003)								I			1 I	1
Brown *et al.* (2010)								I			1 I	1
Amin *et al.* (2007)								I			1 I	1
Thomas *et al.* (2003)								I			1 I	1
Parahoo *et al.* (2000)								I			1 I	1
Poolman *et al.* (2007)								I			1 I	1
Prior *et al.* (2010)								I			1 I	1
Scott *et al.* (2000)								I			1 I	1
Wallen *et al.* (2010)								I			1 I	1
Estrada (2009)								I			1 I	1
Haynes *et al.* (2006)								I			1 I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Van Duppen *et al.* (2007)								I			1 I	1
McKnight (2006)								I			1 I	1
Tilburt *et al.* (2007)								I			1 I	1
Darst *et al.* (2010)								I			1 I	1
Linthorst, Daniels, & Van Westerloo (2007)								I			1I	1
Qian *et al.* (2001)								I			1 I	1
Coffey (2005)								I			1 I	1
Kachuie (2011)								I			1 I	1
Ayre (2009)								I			1 I	1
Waters (2006)								I			1 I	1
Lee (2005)								I			1 I	1
Moyer (2002)								I			1 I	1
Hui (2000)								I			1 I	1
Lee (2000)								I			1 I	1
Suarez-Varela (1999)								I			1 I	1
Kenny (1997)								I			1 I	1
Gill (1996)								I			1 I	1
Crowther (2008)								I			1 I	1
Jemec (2008)								I			1 I	1
Shuval (2007)								I			1 I	1
Lau (2007)								I			1 I	1
Bhatt (2007)								I			1 I	1
Lai (2003)								I			1 I	1
Kingston (2001)								I			1 I	1
Khan (2006)								I			1 I	1

(*Continued*)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Wenban (2003)								I			1 I	1
Hardern (2003)								I			1 I	1
Abeni (2001)								I			1 I	1
Good (2001)								I			1 I	1
Nordin-Johansson (2000)								I			1 I	1
Michaud (1998)								I			1 I	1
Wyatt (1998)								I			1 I	1
Baraldini (1998)								I			1 I	1
Schaafsma (2006)								I			1 I	1
Gabbay (2004)								I			1 I	1
Luker (1992)								I			1 I	1
McCaughan (2005)								I			1 I	1
Cullen (2002)								I			1 I	1
Forsetlund (2003)								I			1 I	1
Smith *et al.* (2000)									I		1 I	1
Green and Ellis (1997)									I		1 I	1
Linzer *et al.* (1988)									I		1 I	1
Landry *et al.* (1994)									I		1 I	1
Villanueva *et al.* (2001)									I		1 I	1
Cabell *et al.* (2001)									I		1 I	1
Langham *et al.* (2002)									I		1 I	1
Stevermer *et al.* (1999)									I		1 I	1
Crowley *et al.* (2003)									I		1 I	1

(Continued)

Summary of 204 unique measures identified across all included systematic reviews	Albarqouni *et al.*[40] (2018)*	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernandez-Dommguez *et al.*[38] (2014)	Glegg and Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)*	Shaneyfelt *et al.*[34] (2006)*	Thomas and Kreptul[35] (2015)	Measures deemed adequate (A) by the authors versus those simply identified (I)	# of systematic reviews having identified the measure
Epling *et al.* (2002)									I		1 I	1
Bell (2000)										I	1 I	1
Bell (2008)										I	1 I	1
Cramer (2001)										I	1 I	1
Dory (2010)										I	1 I	1
Grad (2005)										I	1 I	1
# of measures identified as adequate (A)	6	0	8	0	0	6	0	12	7	0		
# of measures identified (I)	0	35	26	24	16	6	18	70	13	8		
Total	6	35	34	24	16	12	18	82	20	8		

A, to describe a measure deemed "adequate" by the authors of the systematic review; EBM, evidence-based medicine; EBP, evidence-based practice; HCP, health care professional; I, to describe a measure identified by the authors of the systematic review; OT, occupational therapist; OSCE, objective structured clinical examination; PT, physiotherapist.
*The review did not report identifiable information on all included measures (ie, name of measure and/or source study)
Please note that the primary study for each identified measure can be found by retracing the reference in the systematic review(s) that identified the measure.

Appendix VI: Critical appraisal reasoning

Reasoning is provided when a score of 0 or 0.5 is attributed to a criterion (see Table 2).

Albarqouni *et al.*,[40] 2018

- 4: The authors report using Web of Science as the only citation index and a single study as the starting point for the backward-forward citation search; however, they do not mention the reasoning behind either choice (ie, why they did not use more than one citation index; why they did not include more studies that could have led to a more complete yield). Also, only one independent reviewer screened for study eligibility. Further, it is unclear which software was used to manage citations and if any methods were used to minimize errors in searching for studies.
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.

Belita *et al.*,[36] 2020

- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 10: The authors did not provide recommendations for policy and/or practice related to the reported data (ie, the quality of measures).

Buchanan *et al.*,[31] 2016

- 1: The authors did not provide a clear and complete research question or aim (the population is missing and only found later in the methods). It should include the following four key elements: i) the construct of interest; ii) the population(s); iii) the type of measure or instrument(s); and iv) the measurement properties of interest (Prinsen *et al.*,[11] 2018).
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 7: Only one researcher performed extraction without mention of methods to minimize errors in data extraction, such as piloting or verification by another individual. It is not reported whether the rating for quality of instruments was performed by one or more researchers.
- 10 (Half point): The authors make recommendations on eight EBP measures, based on the fact that they have scored at least 3 "excellent" validity markers. Yet, this threshold for making strong recommendations for practice is not supported or explained.

Fernandez-Dominguez *et al.*,[38] 2014

- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 8: The methods used to combine studies are unclear and there are inconsistencies in the methods reported to assess the quality of instruments (ie, different rating scales: "positive, negative, uncertain, absent" vs "clearly reported, partially reported, not reported psychometric properties in primary studies").
- 10: The authors report that the psychometric properties were rated as "positive, negative, uncertain, absent"; however, these results are not reported. Still, the recommendation for practice is that all instruments "were found to be limited as regards the 'constructs' included. Besides, they can all be seen to be lacking as regards comprehensiveness associated to the validation process of the psychometric tests used."[(p.767)] The results are not presented as reported in the methods and the rationale for recommendations is not provided.

Glegg and Holsti,[32] 2010

- 2: In order to be included, measures had to contain *more than one item* designed to evaluate at least one of the two constructs. This restriction in eligibility criteria is stated without rationale and may have led to the exclusion of important tools.

- 3: Only measures containing more than one item for one of the two constructs of interest were included; the authors do not explain why they would exclude one-item per construct and could have excluded important measures. Also, the reasons for excluding articles was not reported.
- 4: Only one reviewer screened for study eligibility.
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 7: Only one researcher performed extraction without mention of methods to minimize errors in data extraction, such as piloting or verification by another individual. It is not reported whether the rating for quality of instruments was performed by one or more researchers.
- 10: Despite the stated research purpose of providing critical appraisal of the literature, the recommendations were superficial and stated, "more research is needed," without further analysis or interpretation of the findings. The authors state, "As no measure's reliability evaluations met the rating criteria for excellent rigor, further reliability testing and the reporting on these results would enhance the utility of these instruments."[(p.230)] Another example attesting to the simplistic nature of the recommendations: "Minor modifications to the most highly ranked measures under evaluation here may provide the most efficient means of generating a comprehensive set of tools for the measurement of knowledge and skills."[(p.230)]

Kumaravel *et al.*,[39] 2020

- 3: The search strategy was not appropriate as they excluded studies that would have contributed to the psychometric evidence of tools as reported in the following citation: "When multiple studies presented the evaluation of the same tool, only the first study which evaluated the psychometric properties of the tool in medical education was included in this review, subsequent studies were considered as duplicates."[(p.3)] It is inaccurate to consider multiple psychometric studies on the same tool as duplicates as per validity theory (ie, validity evidence is ongoing and cumulative).
- 4: Only one reviewer screened for study eligibility at the title and abstract level.
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.

Leung *et al.*,[33] 2014

- 4: Only one reviewer screened for study eligibility at the title and abstract level.
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.

Oude Rengerink *et al.*,[37] 2013

- 1: There is an incomplete purpose statement ("To identify and compare tools to assess Evidence-Based Practice (EBP) behavior among health-care professionals"[(p.129)]) which is lacking the "outcome" or the "the measurement properties of interest" component. Further, in the introduction, the authors report an overview of the methods, which the reader can assume is the research question; however, this is not explicit ("we systematically reviewed the validity, reliability and feasibility of all existing methods to assess EBP behavior of healthcare professionals"[(p.129)]). It should include the following four key elements: i) the construct; ii) the population(s); iii) the type of instrument(s); and iv) the measurement properties of interest (Prinsen *et al.*,[11] 2018).
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 10: The recommendations for EBP instruments are based on the reports of tested validity and reliability in the literature; however, this is not sufficient to claim a recommendation for practice. There is no further analysis or critical appraisal on the quality or rigor of the psychometric information in primary studies. In addition, no further information regarding details on the psychometric evidence is provided.
- 11: The authors report that the Boström questionnaire should be evaluated more extensively and/or existing valid and reliable tools could be combined into an instrument that covers all EBP steps. The

authors state that the Bostrom questionnaire has adequate validity, but do not explain the basis for their statement of "adequate" quality, nor do they explain how and why tools should be combined.

Shaneyfelt *et al.*,[34] 2006

- 1: The research question is incomplete and unclear. In the abstract, the purpose statement reads: "To appraise, summarize, and describe currently available EBP teaching evaluation instruments."(p.1116) There is no mention of the population or the measurement properties of interest. It should include the following four key elements: i) the construct; ii) the population(s); iii) the type of instrument(s); and iv) the measurement properties of interest (Prinsen *et al.*,[11] 2018).
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.

Thomas and Kreptul,[35] 2015

- 1: No research question reported.
- 2: No information on eligibility criteria reported.
- 3: No search strategy provided. Unclear what other methods or "additional sources" were used to identify studies.
- 5 and 6: The authors did not critically appraise included primary studies. As no critical appraisal of included studies was conducted, the process could not have been conducted by two or more reviewers.
- 8: The methods used to combine studies were not reported. It is unclear to the reader why the Fresno test is emphasized (ie, has its own table) in comparison to the other included measures.
- 10: Lack of recommendations for policy and/or practice based on the findings. Recommendations are mainly descriptive and focus on using the Fresno as it has the "best" documentation of validity and reliability; however, further rationale is not provided. Furthermore, overall conclusions relate to the effectiveness of EBM courses and do not describe the tests of EBM as per the research question.
- 11: The last section in the article, "next research steps," does not describe future research avenues as it pertains to measurement of EBM (ie, the topic of the systematic review) but aspects to consider when designing an EBP effectiveness study (updating guidelines/material, including focus groups, including national family medicine organizations to have national EBP competence tests). No ties with findings from the systematic review.

Appendix VII: Additional information on methodological aspects of included reviews

	Albarqouni *et al.*[40] (2018)	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernández-Domínguez *et al.*[38] (2014)	Glegg ánd Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)	Shaneyfelt *et al.*[34] (2006)	Thomas and Kreptul[35] (2015)	Summary
Protocol registered a priori	No	Yes	No	No	No	No	No	No	No	No	1 Yes
Reference to a systematic review methodology	No	Names the Cochrane Handbook for Systematic Reviews for data extraction in the abstract only. No further info on this methodology	Names the Cochrane approach for databases searched in methods section. No further info on this methodology	No	No	No	No	No	No	No	0 Yes, 2 made reference to the Cochrane approach
Reference to a reporting guide line	No	Yes (PRISMA)	No	No	No	Yes (PRISMA in abstract only)	Yes (PRISMA)	Yes (PRISMA)	No	No	4 Yes for the PRISMA guideline (including one that only mentioned it in abstract)
PRISMA flowchart (or equivalent)	Yes	Yes	Yes	Yes (no reference to PRISMA)	No	Yes	Yes (the authors call it the COSMIN framework)	Yes	Yes (no reference to PRISMA)	Yes	9 Yes (including 3 without direct reference to PRISMA)
Included PRISMA checklist (or equivalent)	Yes (in additional files)	No	No	No	No	Yes (in additional files)	No	No	No	No	2 Yes
# of independent reviewers for study eligibility	One (concerns discussed with research team)	Two (discrepancies resolved by discussion)	Two (discrepancies resolved by discussion)	Two (discrepancies resolved by consensus with research team)	One	One (title/abstract level), two (full-text level with discrepancies resolved by consensus)	One (title/abstract level), two (full-text level with discrepancies resolved by consensus between three reviewers)	Two (title/abstract level), different two (full-text level) with discrepancies at any level resolved by consulting a third reviewer	Two (discrepancies resolved by discussion)	Two (discrepancies resolved by discussion)	At least two independent reviewers at all levels of study eligibility (n = 6)
Was detailed syntax for one data base at least available in the publication?	No	Yes	No	Yes	No	Yes	Yes	Yes	No	No	5 reported detailed syntax for one database at least
Were the searches conducted by a library/information specialist?	Yes: mentioned in the acknowledgments (senior information specialist)	Yes: health sciences librarian	NR	NR	NR	Yes: information specialist	NR	NR	NR	NR	3 conducted searches with a library/information specialist
Was reference checking performed?	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	6 performed ref-erence checking of included articles

(Continued)

	Albarqouni *et al.*[40] (2018)	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernández-Domínguez *et al.*[38] (2014)	Glegg ánd Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)	Shaneyfelt *et al.*[34] (2006)	Thomas and Kreptul[35] (2015)	Summary
Reasons for excluding articles reported?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes, but number of studies for each reason unknown	9 provided reasons for article exclusion
Databases searched	Tracked the forward and backward citations (ie, citation analysis technique) of indexed articles from a previous review using Web of Science data	CINAHL, Embase, ERIC, HaPI, MathSciNet, Ovid, hand search relevant journals, gray literature, Canadian health research collection, nursing portals, content experts, conference proceedings (refer to Table 1 in article)	PubMed, EBSCO, and Scopus which includes Africa-Wide Information, CINAHL, ERIC, Health Source: Nursing/academic edition, MEDLINE, PsycARTICLES, and PsycINFO	PubMed, CINAHL, IME, IBECS, LILACS, and Cuiden. In databases of systematic reviews: Cochrane, PEDro and JBI. In journals with summarized evidence: ACP Journal Club, EMB online, EBN online, and Bandolera/Bandolier. In databases of clinical practice guidelines: GuiaSalud, CMA Info-base, National Library Guidelines, and National Guideline Clearinghouse. In the following meta-search engines: EBSCO, BVS, Scopus, Web of Science, Trip Database, and Sumsearch. In evidence-based reviews	Embase, MEDLINE (Ovid), and all Evidence-Based Medicine (EBM) Reviews databases as well as in CINAHL (EBSCO), and PsycINFO databases	MEDLINE, Embase, ERIC, BEME guidelines, Allied and complementary medicine, Cochrane Database of Systematic Reviews (CDSR), and Centre for Reviews and Dissemination (CRD) Databases (Database of Abstracts of Reviews of Effects (DARE).	MEDLINE, Embase, CINAHL, ERIC, Cochrane Database of Systematic Reviews, All EBM reviews, and PsycINFO	MEDLINE (PubMed), Embase (Ovid), the Cochrane Library, CINAHL (EBSCO) and PsycINFO (EBSCO)	MEDLINE, Embase, CINAHL, HAPI, and ERIC databases + table of contents of 8 major medical education journals + EBP websites	MEDLINE, Psy-c INFO, ERIC, the Research & Development Resource Base (University of Toronto), ERC, and CBCA Education. Similar searches were conducted in MEDLINE and other databases.	MEDLINE (n = 8), CINAHL (n = 7), Embase (n = 6), ERIC (n = 6), PsycINFO (n = 5), and the Cochrane Library (n = 4), among others
Time limits	Inception-March 2017	1990-December 2017	Inception-February 2014	1996-September 2013	Varied depending on database (Embase: 1980-present; MEDLINE: 1950-present; CINAHL: 1982-present; PsycINFO: 1887-present) *"present" date is not reported	January 2005-March 2019	1960-April 2013	Inception-July 2011	1980-April 2006	Inception-June 2014	4 from inception

(Continued)

	Albarqouni *et al.*[40] (2018)	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernández-Domínguez *et al.*[38] (2014)	Glegg ánd Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)	Shaneyfelt *et al.*[34] (2006)	Thomas and Kreptul[35] (2015)	Summary
Language limits	No limitations	English	NR	Limited to English, Spanish, French, Italian, and Portu-guese	NR	English	No limitations	No limitations	English	NR	3 NR, 3 no limita-tions, 3 English, 1 includes 5 languages
Study design	Controlled (studies with a separate control group, eg, randomized controlled trials or non-randomized controlled trials), which investigated the effect of EBP educational interventions that aimed to teach at least one component of the main EBP steps (of any format or mode, eg, workshop, course, journal club)	Quantitative or mixed methods study designs; articles that reported findings from the use or psychometric testing of measures that assesses EIDM knowledge, skills, attitudes/ values, and/or behaviors	Descriptive studies that used instruments measuring knowledge, skills, attitudes or behavior related to EBP were included. Mixed methods studies were included if they contained a quantitative component and complied with the aforementioned criteria.	Psychometric studies	Studies referring to the evaluation of knowledge and/or skills of occupational therapists in relation to EBP	Included studies that reported a quantitative and/ or qualitative description of at least one tool used to evaluate EBM in medical education which (a) assessed the dimension(s) of EBM learning, namely reaction to educational experience, attitudes, self-efficacy, knowledge, skills, behaviors, and benefits to patients, (b) assessed different step(s) of EBM, and (c) presented results of the psychometric performance of the tool.	Studies were included in the review if they: (1) recruited nurse clinicians as participants; (2) aimed to measure the EBP knowledge, skills, and/or attitudes of participants; (3) contained a description of the instruments' developmental strategy; (4) presented results of validity or reliability test scores; (5) reported levels of EBP/research utilization knowledge, skills and/or attitudes; and (6) used quantitative methodology. Where a replication study referred to an index study for details of instrument criterion, this index study was also included as it contained information on the developmental strategy of the instrument.	Included original studies that described the development or use of EBP behavior assessment tools	No exclusion based on study design. Included studies that (1) reported an instrument or strategy that evaluated EBP knowledge, skills, attitudes, behaviors, or patient outcomes; (2) contained a sufficient description of the instrument or strategy to permit analysis; and (3) presented results of testing the performance of the instrument or strategy.	NR	N/A
Population	Health professionals (irrespective of the discipline or the level of training: undergraduate, postgraduate, or continuous professional education)	Study sample consists of all nurses or a portion of nurses; conducted in any health care setting	Studies had to include participants who were qualified occupational therapists or occupational therapy students (undergraduate or postgraduate). Studies of rehabilitation professionals were included if occupational therapists were part of the sample.	Studies had to include physiotherapists as the main study population	Occupational therapists	Medical students, medical education	In the context of this review, nurse clinicians are defined as enrolled nurses, registered nurses, and/ or midwives who have completed their basic training and are currently registered to practice in clinical settings.	All health care professionals (ie, physicians, dentists, nurses, and other allied health care professionals, such as physiotherapists, speech-language therapists, occupational therapists, and dental hygienists)	All health care professionals irrespective of level of training	Family medicine and general practice residents	N/A

(*Continued*)

	Albarqouni *et al.*[40] (2018)	Belita *et al.*[36] (2020)	Buchanan *et al.*[31] (2016)	Fernández-Domínguez *et al.*[38] (2014)	Glegg ánd Holsti[32] (2010)	Kumaravel *et al.*[39] (2020)	Leung *et al.*[33] (2014)	Oude Rengerink *et al.*[37] (2013)	Shaneyfelt *et al.*[34] (2006)	Thomas and Kreptul[35] (2015)	Summary
Outcomes	EBP instruments reported in the included studies	Measures that assess EIDM knowledge, skills, attitudes/ values, and/or behaviors with nurses	Survey instruments testing EBP knowledge, skills, attitudes, and behavior with occupational therapists	Any measure intended to evaluate whether practice in physiotherapy is evidence-based, or for measuring all or at least some of the influential barriers and/or facilitators for performing EBPT.	Self-report and competence-based measures to evaluate EBP knowledge and skill. In order to be included, measures had to contain more than one item designed to evaluate at least one of the two constructs.	Objective (non-self-reported) tools that evaluate EBM teaching in medical education	Tools to assess knowledge, skills, and attitudes for EBP	All existing methods to assess EBP behavior	Instrument or strategy that evaluated EBP knowledge, skills, attitudes, behaviors, or patient outcomes	Competence tests of EBM	N/A
Type of measure	Any test or questionnaire	Any test or questionnaire	Any test or questionnaire	Any test or questionnaire	Any test or questionnaire	Only objective (non-self-reported) tools	Any test or ques-tionnaire	Any method of assessing EBP	Any method of assessing EBP	Any test or ques-tionnaire	N/A
Included EBP domain(s)	EBP generally - Did not specify	K, S, A, B	K, S, A, B	EBP generally - Did not specify	K, S	Dimension(s) of EBM learning (reaction to educational experience, A, self-efficacy, K, S, B, and benefits to patients)	K, S, A	B	K, S, A, B, or patient outcomes	EBP generally - Did not specify	N/A

(*Continued*)											
	Albarqouni *et al.*[40] (2018)	**Belita *et al.*[36] (2020)**	**Buchanan *et al.*[31] (2016)**	**Fernández-Domínguez *et al.*[38] (2014)**	**Glegg ánd Holsti[32] (2010)**	**Kumaravel *et al.*[39] (2020)**	**Leung *et al.*[33] (2014)**	**Oude Rengerink *et al.*[37] (2013)**	**Shaneyfelt *et al.*[34] (2006)**	**Thomas and Kreptul[35] (2015)**	**Summary**
Exclusion criteria (if reported)	NR	Studies were excluded if the sample consisted of solely other health care professionals or nursing undergraduate students, or in which data specific to nurses was not reported separately. As well, studies testing or using measures assessing research utilization were excluded.	Systematic reviews, either of instruments measuring these aspects or reviews of EBP surveys, were excluded, but the reference lists of included papers were checked to identify studies that may have been missed. Papers that focused on these aspects in a specific area of practice (eg, stroke) rather than EBP in general were excluded.	Studies where tools were used exclusively for the purpose of performing a situation diagnosis, but in which no reference to design, development, and/or validation process of the tool was made; studies in which purely descriptions regarding the use of research in clinical practice or concerning barriers and/or facilitators for performing EBPT were made; studies in which only a cross-cultural adaptation of an existing specific tool were made, with no further psychometric validation; studies where such instruments were developed and used exclusively in a teaching field and/or for academic research (including those using very recently graduated health professionals).	Those articles describing allied health or health professionals without specific reference to occupational therapists were excluded. Articles describing solely qualitative methodologies were excluded.	Tools which were explicitly designed for use in evaluating EBM teaching for other health care professionals (eg, nurses or dentists). However, if such a tool was later validated for use in medical education, they were included in this review. Qualitative studies discussing perceptions of EBM curriculum that did not satisfy the inclusion criteria, conference abstracts, short notes, comments, editorials, and study protocols were excluded. When multiple studies presented an evaluation of the same tool, only the first study that evaluated the psychometric properties of the tool in medical education was included in this review; subsequent studies were considered as duplicates.	Studies that: (1) used only qualitative methods to explore EBP knowledge, skills, and attitudes (because validity test scores would not be available for comparison); (2) did not mention the background of participants; (3) were abstracts, duplicate, or incomplete reports; and (4) did not report results for each study variable.	Studies about adherence to guidelines, evidence-based care, or quality indicators regarding one particular disease, since these tools address specific behavior regarding the guideline or disease evaluated and outcomes of these studies would likely be hard to extrapolate to other (general) settings. Studies about the evaluation of preclinical students, as they are not working in practice yet. To optimize applicability of the results we excluded randomized controlled trials that evaluate strategies for improving EBP behavior because the evaluation used to assess the strategies may not be feasible outside the trial. Proceedings of conferences were not included as they contained too little information about the assessment methods used.	Studies that reported only satisfaction with a curriculum were excluded.	NR	N/A

A, attitudes; B, behaviors; EBM, evidence-based medicine; EBP, evidence-based practice; EBPT, evidence-based physiotherapy; EIDM, evidence-informed decision making; K, knowledge; N/A, not applicable; NR, not reported; S, skills

Appendix VIII: The breadth and depth of reporting of psychometric properties and measurement characteristics by systematic review

Systematic review; Reporting by measure or study	Psychometric properties									Other characteristics	
		Internal structure			Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness	Interpretability	Feasibility
	Content validity	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance							
Albarqouni *et al.*[40] (2018) By measure	Established (y) when applicable among the six high-quality measures only Proportion of reporting among all measures **Instrument development** section describes the testing population (sample size, level of expertise, area of practice, country) **Instrument description** section: intent of use for some measures	Proportion of reporting "**internal validity**," which includes internal consistency and dimensionality among all measures No further information on dimensionality or factor analysis	Established (y) when applicable among the six high-quality measures only	No specific property reported for this. Mentions available and "validated" translation, but no further information on the cultural validation process (see Berlin questionnaire).	**Interrater reliability** only: Established (y) when applicable among the six high-quality measures only Proportion of reporting among all measures	NR	Proportion of reporting among all measures	**Discriminative validity**: Established (y) when applicable among the six high-quality measures only Proportion of reporting among all measures	**Responsive validity**: Established (y) when applicable among the six high-quality measures only Proportion of reporting among all measures	NR	No specific section for this. In the **instrument description** section, there is information on mode of administration for the six high-quality measures and completion time and scoring time for 1/6 measures.
	No distinction of psychometric evidence by study population when multiple primary studies are referenced for a measure. No actual scores or information on study methods presented for psychometric evidence.										
Belita *et al.*[36] (2020) By measure, then by primary study	Checkmark (y) **Purpose of measure and description**: intent of use described Description of methods to demonstrate content validity (methods, participants, analysis, results)	**Internal structure**: Checkmark (y) *information on factor analysis (factors, results,* data structure, and conclusion)	**Reliability**: Cronbach's alpha scores per measure, per setting (pooling methods unknown)	NR	Test-retest, interrater test scores	NR	"**Relationships to other variables**": Checkmark (y) *A review of the literature was conducted and guiding tables on variable relationships were established. Information on hypothesis, study population, scores* with variation and significance, grouped by type of statistical analysis used (eg, regression, t-test, correlation)			**Acceptability**: information on missing data parameters	**Acceptability**: information on completion time. In **description of measures** section, mode of administration reported.
	This review also includes the validity category "**response process**," which involves understanding the thought processes participants use when responding to items and their consistency with the construct of focus. Checkmark (y) + *overview of study methods, participants, and sample size*										

(Continued)

Systematic review; Reporting by measure or study	Psychometric properties									Other characteristics	
	Content validity	Internal structure			Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness	Interpretability	Feasibility
		Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance							
Buchanan *et al.*[31] (2016) By primary study	**Instrument construction**: information on scale construction and type of administration **Validity**: description of methods to demonstrate content validity: variable detail (methods, results, sample size) Content validity: Graded on a 3-point scale (excellent, fair, poor)	Results with variable detail (structure, sample size, methods, conclusion) Structural validity: graded on a 2-point scale (excellent, poor)	Cronbach's alpha scores with variable detail (raw score with unknown test to study design and subscales to conclusion) Internal consistency: Graded on a 2-point scale (excellent, poor)	Inconsistent detail provided for two translated measures: for Heiwe *et al.*, reference to the source study with unknown language and translation process; for Karlsson and Tornquist, brief mention of the translation process although language unclear Cross-cultural validity: Graded on a 3-point scale (excellent, poor, NA)	Intra-rater reliability and test-retest scores (no information on methods or analysis) **Reliability**: Graded on a 2-point scale (excellent, poor)	NR	For one measure: "no external reference for comparison"(p79)	Variable information without raw data. For one measure: "construct validity: factor analysis revealed 5 factors". For one, "evidence of construct validity". For one, "demonstrated through correlations with other EBP measures" (Table 2) Hypothesis testing: Graded on a 4-point scale (excellent, good, fair, poor)	For one measure, effect size with significance provided and information on outcomes, time, and population	NR	**Clinical utility**: Variable data on clarity of instructions, acceptability and time to complete Clinical utility: graded on a 4-point scale (excellent, good, fair, poor)
Fernaíndez-Domínguez *et al.*[38] (2014) By primary study	Clearly, partially, or not stated **Theoretical ground**: clearly, partially or not stated **In comment section**: intent of use for some measures	NR	Clearly, partially, or not stated Cronbach's alpha scores provided in the discussion only	**In comment section**: reports when two measures are translations with references reported but no further information.		**Reproducibility** (agreement and reliability): clearly, partially, or not stated	Clearly, partially, or not stated	**Construct validity**: clearly, partially, or not stated	Clearly, partially, or not stated	**Floor and ceiling effects**: clearly, partially or not stated	NR
Glegg and Holsti[32] (2010) By measure	**Focus** section: intent of use of measure **Validity** section: variable information on methods for content validity or conclusions **Item selection** section: Graded on 4-point scale (excellent, adequate, poor, N/A)	**Validity** section: For one measure, factor analysis results and brief description of methods	**Validity or reliability** sections: Cronbach's alpha scores with interpretation for two measures	For one measure, reports that it is "translated" with no further information (see Karlsson & Tornquis 2007)	Inter-rater reliability and test-retest scores with variable detail (scores only to scores with CI and interpretation) Graded on a 3-point scale (excellent, adequate, poor)	NR	NR	**Validity** section: Construct validity results with variable detail (from stating "good" to hypothesis, methods, scores, significance)	For one measure, responsiveness score and strength provided without further information on study design	NR	Clinical utility was omitted as "only one tool described such information and as all tools were freely available online or on request from the author and had clear instructions."(p.222)

(Continued)

Systematic review; Reporting by measure or study	Psychometric properties									Other characteristics	
	Content validity	Internal structure			Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness	Interpretability	Feasibility
		Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance							
Kumaravel *et al.*[39] (2020) By measure	Established (y) **Instrument description** section: intent of use for some measures For one measure, brief description of methods to demonstrate content validity	**Psychometric properties** section: For one measure, item response theory results (scores and brief description of methods)	**Internal validity**: Established (y) **Internal reliability (ITC)**: Established (y) **Psychometric properties** section: Cronbach's alpha scores or Spearman's correlation range; Itemtotal correlation scores and method	For one measure, reports that the source questionnaire is in Dutch, but no further information (see the Utrecht Questionnaire (U-CEP)	**Inter-rater reliability**: Established (y) **Psychometric properties** section: scores with variable detail on methods (eg, just score, one measure has descriptive information on the raters)	NR	**Psychometric properties** section: For one measure, criterion validity is said to be established with significance	**Construct validity**: Established (y) **Psychometric properties** section: results with variable detail (some without scores but interpretation of results, some with scores and significance) **Discriminative validity**: Established (y) **Psychometric properties** section: results (variable detail)	**Responsive validity**: Established (y) **Psychometric properties** section: results range from providing no further information to providing vague interpretation of results to detailed information on methods	**Psychometric properties** section: item difficulty range for two measures	**Instrument description** section: type of administration and completion times
	External validity is also included with report on whether it is established (y), however, no further information on this property in the article. It is unclear what the authors mean by "external validity."										
Leung *et al.*[33] (2014) By primary study	**Validity(ies)/ Reliability(ies)**: Content: brief description of methods to demonstrate content validity (no sample size or results; sometimes participants)	NR	**Validity(ies)/ Reliability(ies)**: internal consistency: Cronbach's alpha scores only	Under subheading Content validity: reports "professional translator" for one measure, which was translated into Spanish (EBPQ-19)	**Validity(ies)/ Reliability(ies)**: inter-rater reliability: for one measure, score provided only (no further information)	NR	Authors report searching for this property, but no measures had demonstrated this	**Validity(ies)/ Reliability(ies** Construct validity results only (correlation, % variance, or alpha; no further information)	NR	"Other reported measures like acceptability and interpretability have not been used for validity assessment in this study as there is no apparent way of quantifying them."(p.2187)	
Oude Rengerink *et al.*[37] (2013) By primary study	***Measurement properties****: information on content validity with variable detail (methods, results,* conclusions)	***Measurement properties****: for one measure, reports when factor analysis was tested (no further information)*	***Measurement properties****: variable data from reporting when internal consistency was tested to providing Cronbach's alpha scores*	Mentions when two measures were translations with references reported, but no further information	***Measurement properties****: variable data from reporting when inter-rater or test-retest was tested to providing ICC scores*	NR	NR	***Measurement properties****: Mentions when construct validity was tested (no further information)*	***Measurement properties****: For one measure, reports when responsiveness was tested (no further information)*	NR	**Description of the instrument**: mode of administration sporadically reported
	The only information provided in the paper is through a checklist of whether reliability and/or validity has been tested (only for measures where validity and/or reliability had been tested).										

(*Continued*)

Systematic review; Reporting by measure or study	Psychometric properties									Other characteristics	
		Internal structure			Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness	Interpretability	Feasibility
	Content validity	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance							
Shaneyfelt *et al.*[34] (2006) By measure	Frequency of reporting content validity across all measures. **Validity**: reports when content validity is established for level 1 to 3 instruments (no further information)	**Dimensionality**: frequency of reporting across all measures	Frequency of reporting internal consistency across all measures **Validity**: reports when internal consistency established (yes) for level 1 to 3 instruments (no further info)	NR	**Inter-rater reliability**: scores (Kappa, correlation) for level 1 to 3 instruments	NR	Frequency of reporting criterion validity across all measures **Validity**: reports when criterion validity established for level 1 to 3 instruments (no further information)	Frequency of reporting discriminative validity across all measures **Validity**: reports when discriminative validity established for level 1 to 3 instruments (no further information)	Frequency of reporting responsive validity across all measures **Validity**: reports when responsive validity established for level 1 to 3 instruments (no further information)	NR	**Description**: Mode of administration sporadically reported. Reference to the studies that report time required to administer or score, expertise required for scoring and financial costs (without providing raw data).
Thomas and Kreptul[35] (2015) By primary study	Description of methods to demonstrate content validity: variable detail (methods, sample, conclusions)	NR	**Internal (consistency) reliability**: Cronbach's alpha scores (and subscores) and ITC with varying information on study details	Short description of the translation process for one measure, which was translated into Spanish	**Inter-rater or intrarater reliability**: scores (no further information)	NR	NR	**Concurrent validity**: for one measure, scores provided (correlation with p-value)	NR	NR	**Content validity** section: mode of administration sporadically reported

CI, confidence interval; ICC, intraclass correlation coefficient; ITC, item total correlation; NR, not reported
Information in *italics* signifies that the information is found in an additional file. Bold signifies a different labeling of the category.
Information in underlined text: when specific scores or results of statistical analyses have been explicitly reported by authors.
Unless specified otherwise, the reported psychometric information is presented for all included measures within the systematic review based on availability in primary studies. The psychometric information may not be available for each measure as it may not have been tested or reported in primary studies.

The inclusion and order of psychometric properties and characteristics in the table follows the 10 steps for conducting systematic reviews of patient-reported outcome measures (PROMs) published by the COSMIN group.

Appendix IX: Additional information on the criteria of adequacy at the property level

Buchanan *et al.*,[31] 2016

Rating scale				
Measurement property	**Excellent**	**Good**	**Fair**	**Poor**
Internal consistency: Was an internal consistency statistic calculated for each subscale separately?	Internal consistency statistic calculated for each subscale separately			Internal consistency statistic NOT calculated for each subscale separately
Reliability (intra-rater reliability): Was an ICC (for continuous variables) or a Kappa (for dichotomous/nominal/ ordinal variables) calculated?	ICC or Kappa calculated			No ICC/Pearson/ Spearman correlations calculated. No Kappa calculated–only percentage agreement
Content validity: Was there an assessment of whether all items refer to relevant aspects of the construct being measured?	Assessed if all items refer to relevant aspects of the construct being measured		Poor description of the relevant aspects of the construct being measured	Did not assess if all items refer to the relevant aspects of the construct being measured
Structural validity: Was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed			No exploratory or confirmatory factor analysis performed
Hypothesis testing: Were hypotheses regarding correlations or mean differences formulated a priori?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Cross-cultural validity:				
Were both the original language in which the instrument was developed and the language in which it was translated described?	Both source language and target language described			Source language not known
Did the translators work independently of each other?	Translators worked independently	Assumable that translators worked independently	Unclear whether translators worked independently	Translators did not work independently
Were items translated forward and backward?	Multiple forward and backward translations	Multiple forward but one backward translation	One forward and one backward translation	Only a forward translation
Clinical utility: Were the clarity of the instructions, the format/acceptability of the instrument and the time taken to complete the questionnaire tested?	The clarity of instructions, format/ acceptability of the instrument and time taken to complete were tested	Two of the described aspects of utility were tested	One of the described aspects of utility were tested	No mention was made of any aspects of utility

ICC, intraclass correlation coefficient

Glegg & Holsti,[32] 2010

Appraisal criterion	Focus of the measure	Scale construction	Reliability	Validity	Overall utility
Criterion definition	Primary purpose Population (type of health professional) Evaluation context	Quality of item selection Item weighting Level of measurement	Rigor of standardization studies Reliability Information	Rigor of standardization studies Content validity Construct validity Responsiveness	Overall rating based on scale con- struction and reliability ratings, as well as validity information
Rating or description used in the appraisal	No rating used. Primary purpose defined by original study's author. Measures of perceived knowledge or skill differentiated from objective measures of knowledge or skill.	Rating used for item selection: Excellent: included all relevant characteristics based on comprehensive literature review and survey of experts Adequate: included most relevant characteristics Poor: convenient sample of the characteristics Not applicable: no items address the construct of interest	Rating used: Excellent: more than 2 well-designed studies with adequate (0.60-0.79) to excellent (0.80) reliability Adequate: 1 to 2 well-designed studies with adequate to excellent reliability Poor: no evidence reported, or poor reliability (<0.60) reported	No ranking, as content and construct validity study results may not reflect validity of the tools for the purpose of evaluating EBP knowledge and skills, if they were designed for other primary purposes. Study size and design stated, if applicable. Methods used to increase or demonstrate validity listed	Excellent: excellent scale construction, reliability, and evidence of validity Adequate: adequate to excellent scale construction, reliability, and evidence of validity Poor: poor scale construction, reliability, and/or evidence of validity Includes recommendations for improving overall utility

Leung *et al.*,[33] 2012

Psychometric measures/ levels	Content validity	Construct validity	Criterion validity	Internal consistency	Test-retest reliability	Inter-rater reliability
		Convergent/divergent or discriminant ANOVA (Cohen's f)/ T-Test (Cohen's d) or Eta Squared (n^2)/ Pearson product-moment correlation (r)/Spearman rank-order correlation (ρ) Multitrait-multimethod/Factor analysis: Percentage variance explained and Kaiser-Meyer-Olkin value (KMO) Probability value (*P*) Cronbach's alpha (α)	Concurrent/predictive ANOVA (Cohen s f)/ T-Test (Cohen s d) or Eta Squared/Pearson product-moment correlation (r)/ Spearman rank-order correlation Probability value (*P*) Diagnostic/screening instruments: Area under curve (AUC) Positive likelihood ratio (LR+) Negative likelihood ratio (LR-)	Reliability coefficient (α) Cronbach's alpha Kuder-Richardson 20 (KR-20) / Split-half reliability	Kappa coefficient (K): Landis's K or Fleiss's K Intraclass correlation coefficient (ICC)/Pearson correlation (*r*)/ Probability value (*P*)	Kappa coefficient (K): Landis's K or Fleiss's K Intraclass correlation coefficient (ICC)/Pearson correlation (*r*)/ Probability value (*P*)
A		Cohen's f $\geq$ .40 / Cohen's d $\geq$ 0.80 or $n^2 \geq 0.14$ r or $p = \pm.50$-±1.0 KMO $\geq$ .80 percentage variance $\geq$ 70% $P < 0.05$ $\alpha \geq 0.90$	Cohen's f $\geq$ 0.40 / Cohen's d $\geq$ 0.80 or n^2 $\pm.50$-±1.0 $P < 0.05$ AUC > 0.9 LR+ > 10 or LR- < 0.10	a $\geq$ 0.90	Landis's K $\geq$ 0.81 or Fleiss's K > 0.75 ICC > 0.75 r > 0.95 $P < 0.05$	Landis's K $\geq$ 0.81 or Fleiss's K> 0.75 ICC > 0.75 $r \geq 0.95$ $P < 0.05$

(Continued)						
Psychometric measures/ levels	**Content validity**	**Construct validity**	**Criterion validity**	**Internal consistency**	**Test-retest reliability**	**Inter-rater reliability**
B		Cohen's f = 0.25-0.39 / Cohen's d = 0.50-0.79 or n^2 = 0.06-.13 *r* or *p* = ±0.30-±0.49 KMO = 0.70-0.79 percentage variance ≥ 70% *P* < 0.05 α = 0.80-0.89	Cohen's f = 0.25-0.39 / Cohen's d = 0.50-0.79 or n^2 = 0.06-0.13 r or p = ±0.30-±0.49 *P* <.05 AUC = 0.70-0.90 LR+ = 5.0-10 and LR- = 0.10-0.20	α = 0.80-0.89	Landis's K = 0.61-0.80 or Fleiss's K = 0.60-0.75 ICC = 0.60-0.74 r = 0.90-0.94 *P* < 0.05	Landis's K = 0.61-0.80 or Fleiss's K = 0.60-0.75 ICC = 0.60-0.74 r = 0.90-0.94 *P* < 0.05
C	Expert panel	Cohen's f = 0.10-0.24 / Cohen's d = 0.20-0.49 or n^2 = 0.01-0.05 *r* or *p* = ±0.10-±0.29 KMO = 0.60-0.69 percentage variance ≥ 70% *P* < 0.05 α = 0.70-0.79	Cohen's f = 0.10-0.24/ Cohen's d = 0.20-0.49 or n^2 = 0.01-0.05 *r* or *p* = ±0.10-±0.29 *P* < 0.05 AUC = 0.50-0.69 LR+ = 2.0-5.0 and LR- = 0.50-0.20	α = 0.70-0.79	Landis's K = 0.41-0.60 or Fleiss's K = 0.40-0.59 ICC = 0.40-0.59 *r* = 0.85-0.89 *P* < 0.05	Landis's K = 0.41-0.60 or Fleiss's K = 0.40-0.59 ICC = 0.40-0.59 *r* = 0.85-0.89 *P* < 0.05
D	Group of related clinicians. Feed-back from partici-pants. Literature review	Cohen's f< 0.10 / Cohen's d < 0.20 or n^2 <0.01 *r* or *P* < ±0.10 KMO = 0.5-0.59 percentage variance < 70% P ≥ 0.05 α < 0.69	Cohen's f < 0.10 / Cohen's d < 0.20 or n^2 < 0.01\ *r* or *P* < ±0.10 P > 0.05 AUC ≤ 0.49 LR+ = 1.0-2.0 and LR- = 0.50-1.0	α ≤ 0.69	Landis's K < 0.40 or Fleiss's K < 0.40 ICC < 0.39 *r* ≤ 0.84 or P ≥ 0.05	Landis's K < 0.40 or Fleiss's K < 0.40 ICC ≤ 0.39 r ≤ 0.84 or P ≥ 0.05

Grading of psychometric strength		
Grade of psychometric strength	**Description**	**Example**
Good	Three or more As and/or Bs ± Cor D	A+A+B+C
Adequate	Two As and/or Bs ± Cor D	A+B+C+D
Weak	One A or B ± Cor D	B+D+C
Very weak	One or more C or D only	D+C

Chapter 3.2: Manuscript 2

Challenges and Future Directions in the Measurement of Evidence-based Practice: Qualitative Analysis of Umbrella Review Findings

Jacqueline Roberge-Dao[1,2], Lauren A. Maggio[3], Marco Zaccagnini[1,2], Annie Rochette[2,4], Keiko Shikako[1,2], Jill Boruff[5], Aliki Thomas[1,2,6]

[1] School of Physical and Occupational Therapy, McGill University, Montréal, Canada
[2] Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Canada
[3] Medicine and Health Professions Education, Uniformed Services University, Bethesda, USA
[4] L'École de réadaptation de l'Université de Montréal, Montréal, Canada
[5]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Canada
[6]The Institute of Health Sciences Education, McGill University, Montréal, Canada

Corresponding author: Aliki Thomas, aliki.thomas@mcgill.ca, 3654 Promenade Sir William Osler, Montreal, QC, Canada, H3G-1Y5, 514-398-4496

Published in the *Journal of Evaluation in Clinical Practice* (2022)
http://doi.org/10.1111/jep.13790

Abstract

Objectives: An important aspect in scholarly discussions about evidence-based practice (EBP) is *how EBP is measured*. Given the conceptual and empirical developments in the study of EBP over the last three decades, there is a need to better understand how to best measure EBP in educational and clinical contexts. The aim of this study was to identify and describe the main challenges, recommendations for practice and areas of future research in the measurement of EBP across the health professions as reported by systematic reviews.

Methods: We conducted a secondary analysis of qualitative data obtained in the context of an umbrella review that aimed to compare systematic reviews on EBP measures. Two reviewers independently extracted excerpts from the *results* and *discussion/conclusion* sections of the 10 included systematic reviews that aligned with the three research aims. An iterative six-phase reflexive thematic analysis was conducted.

Results: Challenges reported include limited psychometric testing and validity evidence for existing EBP measures; limitations with the self-report format; lack of construct clarity of EBP measures; inability to capture the complexity of the EBP process and outcomes; and the context-specific nature of EBP measures. Reported recommendations for practice include acknowledging the multidimensionality of EBP; adapting EBP measures to the context and re-examining the validity argument; considering the feasibility and acceptability of measures; and referring to existing frameworks in planning the measurement approach. Areas of future research pertained to the development of comprehensive, multidimensional EBP measures; the need for expert consensus on the operationalization of EBP; more rigorous psychometric evaluations; and the diversification of methodologies to measure EBP.

Conclusion: This study suggests that existing measures may be insufficient in capturing the multidimensional, contextual, and dynamic nature of EBP. There is a need for a clear operationalization of EBP and improved understanding and application of validity theory.

Introduction

Evidence-based practice (EBP) is a core component of health professions education curricula and an expectation for clinical practice worldwide.[1–6] The predominant approach to teaching EBP follows the "five-step model [which] forms the basis for both clinical practice and [education]".[7(p3)] Trainees are taught how to formulate a question (*ask*), search the literature (*access*), *appraise* the evidence, *apply* this information to practice and *assess* outcomes to improve the EBP process in the future.[8–10] Learning outcomes and assessment practices in educational and clinical contexts have largely focused on the knowledge and skill acquisition associated with this step-by-step process.[11]

Despite early uptake of this stepwise approach to EBP, critics have questioned the overreliance on the evidence hierarchy (i.e., valuing higher forms of evidence such as randomized controlled trials (RCTs)) and the idealistic and unrealistic expectation of carrying out these five steps in practice.[7,12–19] Scholars have highlighted challenges associated with the last two steps, *apply* and *assess*, which are largely vague and for which there is little guidance on how to integrate empirical findings with patient preferences, clinical expertise and contextual factors (e.g. the lack of material resources into decision-making).[12,20] Still, the 2018 consensus statement identified 68 core competencies to include in EBP teaching and rated "practicing the five steps of EBP" as highest on their scale of relative importance of core competencies.[21]

There has been no shortage of discussion on the philosophical, methodological, and practical underpinnings of EBP.[1,3,7,10,12–14,16,22–30] These discussions have led to an evolution in how EBP is conceptualized.[31] For instance, there is greater recognition that evidence-based decision-making is not only a complex context-specific process but that it is greatly influenced by systems, organizations, and clinician as well as patient beliefs.[32,33] The question "what constitutes evidence" has been the topic of much deliberation, with a subtle shift towards a pluralistic and more inclusive view of evidence, and an acceptance of methodologies other than RCTs (e.g., observational, case studies).[15,16,29,31,33–38] While these shifts may better cater to the multifaceted nature of the clinical decision-making process, there is no consensus on how we can best teach, practice and measure EBP.[39] Recently, scholars have considered the need for a renewed collective reflection and consensus to support a more contemporary conceptualization of EBP.[31]

Although the teaching and practice of EBP have been central to the conversations, the measurement of EBP has become an increasingly contentious issue. Measurement is a foundational aspect of both research and educational processes and can reflect important beliefs on what is important to measure.[40] EBP measures are used to determine if future clinicians have acquired EBP-related competencies and if they continue to apply the process in practice.[41] Robust measurement of EBP is warranted to determine if teaching strategies and practice interventions are achieving targeted outcomes and if the resources being allocated in clinical and academic spheres are justified. Given that EBP is a professional requirement,[4–6] the interpretation of data from these measures can have important consequences. It is critical that EBP measures be developed based on best measurement practices and provide accurate, relevant, and meaningful information to those who use them. Despite the importance of sound measurement of this complex and multidimensional process, guidance for measuring EBP is sparse.[31]

Our team conducted an umbrella review on systematic reviews (SRs) of psychometric properties of EBP measures.[42] Specifically, we estimated the quality of SRs on EBP measures across healthcare professions and identified differences between reviews regarding the recommended measures and the approaches used to assess the adequacy of EBP measures. The quality of the 10 included SRs was low; none of the SRs assessed the quality of primary studies or adhered to SR methodological guidelines. We identified 204 unique EBP measures from various healthcare professional groups (e.g., nursing, rehabilitation). Of the 204, authors of the SRs identified only 27 measures as adequate. Multiple approaches were used to assess the adequacy of measures (e.g., the COSMIN checklist,[43] the Standards of Psychological and Educational Testing[44]).

Considering the fragmented measurement practices, there is a need to go beyond a descriptive comparison of SRs of psychometric properties, to more deeply understand the common points of tension related to measuring EBP in educational and clinical contexts. The aim of this paper is to identify and describe the main challenges, recommendations for practice and areas of future research in the measurement of EBP as reported in SRs of EBP measures.

Methods

Methodology

This was a secondary data analysis of qualitative data obtained in the context of an umbrella review conducted following the Joanna Briggs Institute methodology for umbrella reviews and according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines."[45,46] We searched five databases MEDLINE (Ovid), Embase (Ovid), CINAHL (EBSCO), PsycINFO (Ovid), and ERIC (EBSCO) from inception to January 18, 2021. We included SRs in English or French that identified measures that assessed EBP as a whole or of constituent parts. A 'measure' was defined as 'a method to collect information' and includes questionnaires, instruments, and other measurement methods. Complete information on the search and screening methods, analysis and the assessment of methodological quality of included SRs is published elsewhere.[42]

Data extraction

Two reviewers independently extracted excerpts from the *results* and *discussion/conclusion* sections of each SRs that aligned with the three research aims: i) reported challenges, ii) recommendations for practice, and iii) areas of future research. The reviewers compared the extractions and discussed any discrepancies to reach consensus.

Data analysis

We used reflexive thematic analysis to inductively identify and describe common themes across the data extracted.[47,48] Our analysis involved six iterative phases: 1) data familiarization, 2) initial code generation, 3) generating (initial) themes, 4) reviewing themes , 5) theme defining and naming and 6) report production.[47,49] Themes were developed at the semantic level whereby the surface meaning (i.e., the explicit content) of the data were captured. Following the familiarization with the excerpts, the first author (JRD) generated initial codes to describe the data while keeping the research questions in mind. A second reviewer (MZ) reviewed all the codes and provided feedback on the relevance of codes and potential for new codes. Based on the feedback, JRD refined the codes to improve coherence between codes, excerpts, and the research question. JRD sorted and combined codes into initial themes using a mind-map to make visual links between similar and connecting concepts.[47] Then, JRD and AT discussed and reviewed the themes to identify any overlapping ideas. JRD iteratively reviewed the excerpts and themes to ensure a "coherent story" [47(p22)] and attributed names and descriptions to each theme.

The entire research team periodically reviewed the themes alongside the excerpts and provided feedback on the comprehensibility and accuracy of the analysis. Analysis continued until all co-authors were satisfied with the themes and descriptions, and no new information was deemed relevant.

This approach to thematic analysis acknowledges that the researcher's subjectivity is an important part of the analytic process. Themes are actively developed by researchers who have backgrounds, interests and beliefs which may influence the organization and reporting of the data.[50] JRD is an occupational therapist (OT) and doctoral candidate with research interests pertaining to the measurement of EBP rooted in modern measurement theory. Co-authors include university researchers and professors with expertise in EBP (AT, LM, AR), knowledge translation (AT, AR, KS), and medical education (AT, LM), researchers with expertise in qualitative methods (AT, LM, AR, KS), OTs (AT, AR, KS), a respiratory therapist and doctoral candidate with research experience in EBP (MZ), and health sciences librarians and researchers (LM, JB).

The definitions attributed to key terms may also influence how research is conducted.[51] In this text, we define a *construct* as a concept to explain a phenomenon not directly observable or measurable, *conceptualization* as "the process whereby concepts are defined theoretically", and *operationalization* "refers to the operations or procedures needed to measure the concept(s)" whereby "variables rather than concepts are the focus." [52(pp162-165)]

Results

Included systematic reviews

The umbrella review search yielded 3572 articles and 10 were included (See Table 1 for an overview).

Table 1: Description of included systematic reviews presented in alphabetical order of first author (n=10)

SR ID# (publication date)	**Article title**	**Population**	**EBP context**	**EBP domains**

SR 1- Albarqouni, Hoffmann & Glasziou (2018)	Evidence-based practice educational intervention studies: a systematic review of what is taught and how it is measured	All healthcare professionals	Clinical practice, education	EBP
SR 2- Belita et al. (2020)	Measures of evidence-informed decision-making competence attributes: a psychometric systematic review	Nurses	Clinical practice	EIDM knowledge, skills, attitudes, behaviors
SR 3- Buchanan, Siegfried & Jelsma (2016)	Survey Instruments for Knowledge, Skills, Attitudes and Behavior Related to Evidence-based Practice in Occupational Therapy: A Systematic Review	Occupational therapists	Clinical practice, education	EBP knowledge, skills, attitudes, behaviors
SR 4- Fernández-Domínguez et al. (2014)	Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the literature	Physical therapists	Clinical practice	EBP
SR 5- Glegg and Holsti (2010)	Measures of knowledge and skills for evidence-based practice: A systematic review	Occupational therapists	Clinical practice	EBP knowledge, skills
SR 6- Kumaravel et al. (2020)	A systematic review and taxonomy of tools for evaluating evidence-based medicine teaching in medical education	Medical students	Education	EBM
SR 7- Leung et al. (2014)	Systematic review of instruments for measuring	Nurses	Clinical practice	EBP knowledge,

	nurses' knowledge, skills and attitudes for evidence-based practice			skills, attitudes
SR 8- Oude Rengerink et al. (2013)	Tools to assess Evidence-Based Practice behavior among healthcare professionals	All healthcare professionals	Clinical practice	EBP behaviors
SR 9- Shaneyfelt et al. (2006)	Instruments for evaluating education in evidence-based practice. A systematic review	All healthcare professionals	Clinical practice, education	EBP knowledge, skills, attitudes, behaviors
SR 10- Thomas and Kreptul (2015)	Systematic Review of Evidence- Based Medicine Tests for Family Physician Residents	Family medicine residents	Education	EBM
SR, Systematic review; EBP, Evidence-based practice; EIDM, Evidence-informed decision-making; EBM, Evidence-based medicine				

Thematic analysis findings

We analyzed excerpts from the 10 SRs and produced five themes describing the main challenges associated with measuring EBP, four main recommendations for practice, and four areas of future research (See Table 2 for overview of themes).

Table 2: Overview of the main themes from excerpts of systematic reviews on EBP measures

CHALLENGES
Limited psychometric testing and evidence for existing EBP measures (SRs[‡] 1, 3, 4, 5, 7)
Limitations with the self-report format of most EBP measures (SRs 3, 5, 7, 9)
The lack of construct clarity of EBP measures (SRs 3, 4, 5, 7, 8)
Inability to capture the complexity of the EBP process and outcomes (SRs 1, 2, 6, 8, 9, 10)
EBP measures are context-specific (SRs 2, 3, 4, 7, 8, 10)

RECOMMENDATIONS FOR PRACTICE
Acknowledge the multidimensionality of EBP (SRs 2, 6, 8)
Consider the feasibility and acceptability of measures (SRs 1, 2, 8, 9)
Adapt EBP measures to your context and re-assess preliminary psychometric properties (SRs 1, 2, 3, 5, 6, 7)
Refer to existing EBP frameworks and guidelines in planning the assessment approach (SRs 1, 6, 9)
FUTURE RESEARCH
Focus on developing comprehensive rather than unidimensional EBP measures (SRs 1, 2, 4, 5, 6, 8)
Expert-level consensus is warranted regarding the measurement EBP (SRs 1, 3, 4, 5)
More rigorous psychometric evaluations are warranted (SRs 2, 5, 6, 8)
Diversify the methodology used to assess EBP (SRs 1, 5, 6, 7, 8, 10)
EBP, Evidence-based practice; SRs, Systematic reviews

Challenges

Five reviews identified that existing EBP measures lack robust and comprehensive psychometric evidence: “in general, there was limited consideration of measurement properties. Eighteen instruments reported no validity and reliability testing at all.” (SR3) Only a few measures, if any, were reported as having strong and high-quality psychometric evidence. Few measures were associated with more than one source or type of validity (e.g., only concurrent validity was examined). A minority of primary studies, as reported by three SRs, (SRs 2, 6, 9) considered the feasibility or acceptability of measures despite these aspects “being identified as a critical aspect of psychometric assessment”. (SR2)

Although self-report is the most frequently used format for EBP measures, it is considered “susceptible to recall and social desirability biases.” (SR7) “[R]espondents may overestimate their involvement to portray themselves in a positive way, thereby providing inaccurate information.” (SR3)

The definition of EBP varies across studies: “sometimes EBP is defined as the adherence to guidelines, whereas others regard EBP as the integration of all components of EBP in clinical practice.” (SR8) There are variations and inconsistencies regarding how EBP constructs are defined and operationalized across primary studies: “a variety of concepts were used for EBP learner outcomes, researchers failed to define the constructs measured and varying definitions were used for EBP learner outcomes." (SR3) Authors reported differences with the scope of included items and level of detail of items for the same construct. Alternatively**,** "identical items (in terms of their formulation) were found in different factors or constructs." (SR 4) The lack of reference to a theoretical model in the development of EBP measures was said to explain, "to a large extent", "all of these difficulties and shortcomings found upon analysing the instruments." (SR 4)

Three SRs (SRs 2, 6, 8) revealed that the step-wise approach to measuring EBP does not account for the intricacies involved in integrating the factors for decision-making as per the three-circle model of EBP: "EBM is not just about the ability to ask the right question, followed by searching and appraising the quality of evidence. It is bringing together clinical expertise, patient values and current best evidence into clinical decision making." (SR 6) Authors challenged the assumption that completing a series of tasks or steps equates to EBP competency, "[current measures] do not conceptually reflect an assessment of competence, defined as quality of ability or performance to an expected standard but rather, focus on mere completion or frequency of completing tasks." (SR2)

Furthermore, patient-related EBP outcomes were rarely, if at all, measured. The impact of EBP "is often latent and distant" rendering it difficult to isolate "the impact of educational interventions" from other influences such as "the dominant role of the overarching team and health care system." (SR1)

EBP measures are developed for use in specific contexts; as such, they cannot be separated from the original context without important considerations for the applicability of items and relevance of the psychometric evidence. Measures are not necessarily transferable to other contexts due to the potential irrelevance of items and changes in the interpretation of results when changing settings: **"**the differences between [medicine and nursing], and physiotherapy were not taken into account." (SR4) Nonetheless, authors of primary studies used these measures in other contexts (without re-examining the psychometric evidence) with the assumption that the psychometric properties would transfer.

Recommendations for practice

The multidimensionality of EBP as it pertains to "the inextricable link between knowledge, skills, attitudes and behaviours to comprise professional competence" (SR2) must be acknowledged in measurement practices. It is recommended to avoid focusing on single EBP domains (e.g., knowledge, attitudes) because "enacting EBP behaviour without the proper knowledge and skills can lead to wrong conclusions." (SR8) EBP is influenced by numerous interrelated factors which must all be assessed for a comprehensive and holistic understanding of EBP.

The feasibility and acceptability of measures should be considered to ensure that the selected measure aligns with existing resources and circumstances and to avoid attrition and respondent fatigue. It is important to "[take] into account tool completion time given demands of busy clinical environments and if high rates of missing data > 10% are present." (SR2)

Psychometric evidence is specific to the context (population, setting and purpose of assessment) in which the original study took place. When using a measure, "modifications to item terminology may [be] required to adapt the reviewed tools for use with [healthcare professionals] in different practice settings, countries, or cultures." (SR5) Psychometric properties must be re-assessed if the participants (profession, level of expertise), context and purpose differ from the original study or if the measure has been modified from the original version.

When planning to measure EBP, authors suggest referring to existing EBP frameworks and guidelines[53,54] such as the Consensus Statement outlining the 68 core competencies[21] which "may help to harmonise the content of EBP educational interventions, and with possibly flow-on effect to the measured outcomes." (SR1)

Future research

There is a need to develop a comprehensive measure "which can evaluate all steps of EBM and educational outcome domains" (SR6) and can address the multifaceted and contextually dependent nature of EBP. When selecting a measurement approach, it is important to consider "the organizational aspects of the different practice settings that are characteristic of physical therapists" (SR4) and **"**other aspects of critical appraisal, including demonstrating an understanding of the client and context and how this information influences the evidence decision-making process…" (SR5)

Expert-level consensus is warranted on the conceptualization and operationalization of EBP. The development of measures should be "based on an agreed theoretical construct of evidence-based [practice]" (SR4) and "a standard set of definitions of EBP learner attributes should be devised." (SR3) Additionally, consensus is needed for the "development and agreement on a core set of valid and reliable recommended instruments to measure outcome domains." (SR1)

Authors called for "more rigorous and consistent psychometric testing" (SR2) of EBP measures "to improve the utility of existing and future measures." (SR5) "Researchers also need to publish information on the feasibility of implementing the tools—time taken to complete and grade along with any other resource implications." (SR6)

Broadening the methodologies used (i.e., study design and methods) to assess EBP may allow for a richer and more nuanced understanding of EBP. Indeed, "interviews and observations may give more in-depth information." (SR8) Some suggestions from the articles included focus groups, (SR2) audit, (SR6) activity diaries, (SR6) objective structured clinical examination, (SR1) and audio-recording in clinics. (SR1)

Discussion

Our findings reveal widespread challenges regarding the measurement of EBP and highlight reported recommendations for research and practice in addressing some of these challenges.

Unclear and varied conceptualizations of EBP

A clearly defined construct is fundamental to the development and utility of any measure, as the absence of one affects the integrity of the measure and the validity of the inferences made.[55–57] Conceptualizations of EBP vary across primary studies which compounds the confusion regarding what EBP really "is" and consequently, what is important to measure.[52] This variability in definitions across primary studies has an impact at two levels: first, it results in discrepancies at the construct-level whereby the conceptualization of EBP may or may not include certain constructs (e.g., skills for critical appraisal, adherence to guidelines, the full five-step EBP process). Second, there are differences in the inclusion and level of detail of items assessing the same construct, a feature referred to as "homonymy" [58] where different items are included in measures said to represent the same construct. These measurement inconsistencies reflect linguistic and conceptual ambiguity and may increase the strain on the individual who is choosing an appropriate measure for use.[59] Consistent with findings in implementation science, fuzzy or poorly defined constructs may lead to the misuse of measures, perpetuate discrepancies across future studies and threaten the developing knowledge base. [58,59] As suggested by Thomas et al.,[31] our results highlight a pressing need for expert-level consensus regarding the conceptualization and operationalization of EBP, which would advance the measurement of EBP.

A connection likely exists between the lack of a clearly defined construct, the absence of theoretical grounding in the development of EBP measures, and a need to develop more coherent EBP theory. [1] Given the reciprocal relationship between theory and measurement, [59] the lack of referencing to theoretical frameworks (as identified in this study) may lead to inconsistencies in how EBP is operationalized. A clear theoretical articulation of the mechanisms of interest is required to support the inferences and conclusions made by EBP measures in education, research, and practice. As Nilsen reports, "a 'good theory' provides a clear explanation of how and why specific relationships lead to specific events". [60(p20)] Despite the numerous models describing EBP components and steps (e.g., three-circle framework[61] and the "five-step process model" [7]), and as Djulbegovic and Guyatt have argued,[1(p1)] there is still a need to develop more coherent theory of EBP as a health-care decision-making process.[31,58,62]

The complexity of EBP as a multidimensional and dynamic phenomenon

Our findings reveal a tension with the widespread assumption[31] that performing the five EBP steps corresponds to "doing EBP". Our findings suggest that the five-step operationalization of EBP cannot adequately capture the multidimensionality and unpredictability of EBP that exists in the clinical context. As such, an important distinction must be made on the inferences that are being drawn; *ability* to complete the five steps and the *actual integration* of evidence into clinical decision-making are different concepts. A clinician's actual integration of evidence into decision-making can be influenced by personal (e.g., self-efficacy in integrating research findings) and external factors (e.g., material resources, patient values), in addition to the individual's pre-requisite ability to do so.[63]

These results raise questions regarding current measurement practices using the five steps to assess ones' ability to enact EBP. First, how can we adequately measure one's ability to *apply* or *evaluate* (steps 4 and 5) when the operationalization of these steps is vague and oversimplified?[12,20] For example, how would evaluators operationalize competence in "integrating [patient] preferences"[21(p8)] or in explaining "different strategies to manage uncertainty in clinical decision making"? [21(p8)] Second, does measuring competency with the five steps effectively reflect the subtleties of evidence-based decision-making when faced with uncertainties and limited resources? Third, are these measures being used because they assess what is meaningful or because they are frequently used and readily available?

While measuring singular dimensions (e.g., knowledge) or competency with the five steps may be warranted in some didactic contexts, we suggest being critical when interpreting results from these measures. Our findings suggest that the multidimensional nature of EBP must be acknowledged in future measurement practices. EBP does not have an immutable or fixed quality. Clinician behavior and the extent to which they will integrate evidence into decision-making varies with each clinical encounter due to multiple interacting parts such as the nature and strength of the evidence, patient heterogeneity, time and resources available and skills of the clinician.[64] The dynamic and synergistic interplay between individual, organisational, and system-level factors that influence whether a clinician will integrate evidence is overlooked in current measurement approaches to EBP.

We invite users of EBP measures to consider whether the five-step process successfully embodies the intended ethos, value, and practice of being evidence-based and whether operationalizing EBP through the five steps will result in the improvements in quality of care that EBP was always intended to accomplish.

Although improving health outcomes is one goal of EBP and "benefits to patients" is outlined as an assessment category in the Sicily Statement, [53] our results point to an absence of patient-related outcomes. This has been well documented and includes critiques of the evidence base for EBP's.[13,66–69] It is challenging to isolate the impact of EBP when other factors may influence health outcomes (e.g., clinician-related, patient-related, organizational, or sociopolitical). Still, given the assumption that EBP leads to improved patient outcomes, patient outcomes must be incorporated into the design of future EBP studies.

The advantages of measuring patient outcomes in the context of EBP are multi-fold. First, this data will contribute to the knowledge base on the supposed causation between EBP and improved patient outcomes. Second, a more explicit examination of the association between patient outcomes and clinician behavior may validate the need to remain cognizant and critical of the impact of one's clinical decisions. Third, clinicians' attitudes towards EBP may be more favourable if patient outcomes improve following the integration of evidence.[70] Lastly, including patient outcomes may raise awareness on the importance of EBP within the patient population;

patients may be more forthright in their expectations for evidence-based care which may foster greater clinician accountability for the integration of evidence in decision-making.

Context-specific nature of EBP measures

The context-specificity of EBP measures reduces their transferability and comparability of findings across professional groups, settings, and purposes of assessment. Our findings reinforce the urgency of adapting EBP measures to the context where it is being applied (i.e., determining the applicability of each item) and reassessing the psychometric evidence before interpreting results and drawing conclusions. Once measures are adapted to better reflect the context, users must explicitly report on how the measure was adapted and how the modifications affect the interpretation of the data.[59,71]

Several authors of primary studies applied EBP measures in contexts different than those of the original context and assumed that the psychometric evidence would carry over. Though delving into the reasons why this is the case is beyond the scope of this paper, we propose that it may be due to: first, users of EBP measures may have limited psychometric literacy and resources, such as time to re-evaluate the psychometric evidence; and/or second, there is a widespread understanding of validity as a property that is intrinsic to the measure, regardless of with whom, where, why or how it is used (this conceptualization is often illustrated when one refers to a "validated measure"). All but one SR[72] conceptualized validity as a test characteristic, where validity is seen as a fixed quality, which could be attributed to conscious misuse, lack of awareness[73] or to the pragmatic nature of off-the-shelf solutions in a resource-strained system.[74(p858)] Our findings suggest that there may be confusion regarding the various conceptualizations of validity[74] and applications to practice, we encourage readers to learn about the different validity discourses, methods and implications.[75]

Psychometrics in EBP measurement

Limited psychometric testing and evidence for EBP measures may be partly attributed to the measure being developed for the specific assessment context; this can lead to a one-time use phenomenon.[58] Though authors of included SRs call for more rigorous psychometric evaluations, they fail to explain this statement or provide reference to methodological guidance. These findings warrant questions such as "what is a *more rigorous* psychometric evaluation?" and "what is considered *strong* evidence?" The criteria used to determine the strength of psychometric evidence and the adequacy of measures varied for each SR and reflects authors' understanding of validity.[42] *The Standards of Psychological and Educational Testing* which are considered the "gold standard" for assessment practices in the social sciences (e.g., education, psychology, and employment) were only referenced by two SRs[41,72] (although only one[72] adequately applied the methods). *The Standards of Psychological and Educational Testing* conceptualize validity as a unitary concept and describe validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests." [76]

Nine of the 10 included SRs opted for an approach to validity that is more common in the clinical sciences such as the COSMIN approach for patient-reported outcome measures.[77] This approach categorizes validity into three (sometimes four) types of validity and tends to promote the conceptualization of validity as a static property of the measure.[78] The conceptualization of validity and the application of validity theory is the object of dispute in many fields.[79,80] It is crucial that authors of measurement studies 1) adhere to a specific validity framework and justify their perspective; 2) describe the argument for why the measure is appropriate for the population, context and purpose of assessment; and 3) ensure coherence in the application of the methodology, in the interpretation and reporting of results.[59,74] We invite experts in the field of measurement to reflect on and discuss the relevance and applicability of validity theory to the measurement of EBP.

Methods and methodology for measuring EBP

Authors of four SRs emphasize the recall and social desirability biases associated with self-report questionnaires and discourage their use while promoting “objective measures”. We recognize that self-report may introduce a social desirability bias for certain constructs, notably attitudes towards EBP. As EBP is a well-known and widely accepted professional competency, we do not expect learners or clinicians to be opposed to it. However, the self-report format can be useful for the measurement of some constructs given the resource-constrained reality of clinical settings; self-report may be a pragmatic option for those lacking time or human resources to conduct an extensive performance-based evaluation.[81] Indeed, in the related field of implementation science, scholars agree that although self-report is prone to bias, using “self-report makes good sense given that many salient constructs pertain to perceptions of individuals involved.”[59(p5)] Constructs such as self-efficacy[82] in integrating evidence or the perception of available resources to facilitate EBP can only be measured using self-report. These self-reported constructs are, in our opinion, vitally important to measure to preserve the clinician's expertise and agency as the penultimate influence in clinical decision-making.[83]

We question this proclivity towards "objective" measures and invite reflection on the underlying epistemological assumptions and the biases at work in favoring these methods.[84] Adhering to the notion of "objectivity" reflects a broader empiricist viewpoint often debated in EBP which "[emphasizes] the systematic, reproducible, unbiased observations" and assumes the superiority of certain methods over others.[64]

Our findings suggest a need for diversification of methodological approaches and the use of qualitative methodologies to allow for a richer understanding of EBP. As argued elsewhere, [31,69] we encourage a multi-method approach to EBP measurement to allow for triangulation.[85] When introducing qualitative data, one must consider the method of synthesizing the information, the method of aggregating it with other forms of data and the considerable amount of time and resources needed.[59,86]

We encourage those interested in the measurement of EBP to worry less about the "objectivity" of measures and more about the quality of the items in measuring variables of interest, the appropriateness of measurement data in the context of study and the validity of inferences being made for a specific purpose with real-world consequences. Attention should be given to conceptual models underlying measurement scales (i.e., unidimensionality), intrinsic technical limitations of measures and the appropriateness of summing items of different constructs.[87]

Limitations

Our analysis only considered excerpts from included reviews at the semantic level whereby the surface meaning was used to develop themes; other potentially descriptive factors such as choice of words were not collected which may have led to rich findings. The scope of our findings was limited by the amount of detail provided in the 10 SRs and in some instances, elaboration of concepts was scant. As we deductively extracted excerpts for each research question, there is a potential loss of context and granularity that is found across paragraphs due to the interconnectedness between challenges, recommendations for practice and areas of future research.

Conclusion

This study extends our knowledge on EBP measurement across the health professions and provides insights on how to better develop, select, and use measures to assess to complex process of EBP. Our findings suggest that EBP constructs are often ill-defined, which may lead to confusion on what is being measured, possible misuse of measures and threats to the validity of inferences being made. We highlight the interdependence of theory and measurement, the contextual and dynamic nature of EBP and the need for a clear conceptualization of EBP that acknowledges uncertainty and the complex clinical decision-making process when faced with equally important yet different forms of knowledge. We discuss the practicality of current EBP operationalizations and invite readers to reflect upon their own assumptions regarding how EBP is being measured. It is our hope that ongoing discussions of this review will serve to advance the study and practice of EBP such that researchers, educators, clinicians, and patients can all benefit and find meaning in the data from EBP measures.

References

1. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *The Lancet*. 2017;390(10092):415-423. doi:10.1016/S0140-6736(16)31592-6

2. Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. *Acad Med*. 2013;88(8):1088-1094.

3. Guyatt GH. Evidence-based medicine. *ACP J Club*. Published online 1991:A-16.

4. Frank JR, Snell L, Sherbino J, Boucher A. CanMEDS 2015. *Physician Competency Framew Ser I*. Published online 2015.

5. CAOT, ACOTRO, ACOTUP, PAC. Joint position statement on evidence-based occupational therapy. *Can J Occup Ther*. 2009;66(5):267-269. https://www.caot.ca/document/3697/J%20-%20Joint%20Position%20Statement%20on%20Evidence%20based%20OT.pdf

6. CPA. Position statement on evidence-informed practice in physiotherapy. Published online 2012. https://physiotherapy.ca/sites/default/files/positionstatements/evidence-informed-practice-position-statement-english.pdf

7. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1

8. Ghali WA, Saitz R, Sargious PM, Hershman WY. Evidence-based medicine and the real world: understanding the controversy. *J Eval Clin Pract*. 1999;5(2):133-138. doi:10.1046/j.1365-2753.1999.00183.x

9. Sackett DL. Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest*. 1986;89(2, Supplement):2S-3S. doi:10.1378/chest.89.2_Supplement.2S

10. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *Br Med J*. Published online 1996:71-72.

11. Bhandari M, Giannoudis PV. Evidence-based medicine: What it is and what it is not. *Injury*. 2006;37(4):302-306. doi:10.1016/j.injury.2006.01.034

12. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014

13. Miles A, Bentley P, Polychronis A, Grey J. Evidence-based medicine: why all the fuss? This is why. *J Eval Clin Pract*. 1997;3(2):83-86. doi:10.1046/j.1365-2753.1997.00103.x

14. Miles A, Bentley P, Polychronis A, Grey J, Price N. Recent progress in health services research: on the need for evidence-based debate. *J Eval Clin Pract*. 1998;4(4):257-265. doi:10.1111/j.1365-2753.1998.tb00084.x

15. Miles A, Grey JE, Polychronis A, Price N, Melchiorri C. Developments in the evidence-based health care debate - 2004: Evidence-based health care debate. *J Eval Clin Pract*. 2004;10(2):129-142. doi:10.1111/j.1365-2753.2004.00501.x

16. Norman GR. Examining the assumptions of evidence-based medicine: Examining the assumptions of EBM. *J Eval Clin Pract*. 1999;5(2):139-147. doi:10.1046/j.1365-2753.1999.00197.x

17. Rogers W, Hutchison K. Evidence-Based Medicine in Theory and Practice: Epistemological and Normative Issues. In: Schramme T, Edwards S, eds. *Handbook of the Philosophy of Medicine*. Springer Netherlands; 2017:851-872. doi:10.1007/978-94-017-8688-1_40

18. Tikkinen KA, Guyatt GH. Understanding of research results, evidence summaries and their applicability—not critical appraisal—are core skills of medical curriculum. *BMJ Evid-Based Med*. Published online 2021.

19. Vandenbroucke JP. Observational Research and Evidence-Based Medicine: What Should We Teach Young Physicians? *J Clin Epidemiol*. 1998;51(6):467-472.

20. Barratt A. Evidence based medicine and shared decision making: the challenge of getting both evidence and preferences into health care. *Patient Educ Couns*. 2008;73(3):407-412.

21. Albarqouni L, Hoffmann T, Straus S, et al. Core competencies in evidence-based practice for health professionals: consensus statement based on a systematic review and Delphi survey. *JAMA Netw Open*. 2018;1(2):e180281-e180281.

22. Ashcroft RE. Current epistemological problems in evidence based medicine. *J Med Ethics*. 2004;30(2):131-135. doi:10.1136/jme.2003.007039

23. Couto JS. Evidence-based medicine: a Kuhnian perspective of a transvestite non-theory. *J Eval Clin Pract*. 1998;4(4):267-275. doi:10.1111/j.1365-2753.1998.tb00085.x

24. EBM working group. Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama*. 1992;268(17):2420.

25. Grahame-Smith D. Evidence based medicine: Socratic dissent. *BMJ*. 1995;310(6987):1126.

26. Greenhalgh T, Howick J, Maskrey N, for the Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348(jun13 4):g3725-g3725. doi:10.1136/bmj.g3725

27. Greenhalgh T, Ozbilgin MF, Prainsack B, Shaw S. Moral entrepreneurship, the power-knowledge nexus, and the Cochrane “crisis.” *J Eval Clin Pract*. 2019;25(5):717-725. doi:10.1111/jep.13124

28. Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. *BMC Health Serv Res*. 2003;3(1):14. doi:10.1186/1472-6963-3-14

29. Tonelli M. The philosophical limits of evidence-based medicine. Published online 1998.

30. Upshur REG. Seven characteristics of medical evidence. *J Eval Clin Pract*. 2000;6(2):93-97. doi:10.1046/j.1365-2753.2000.00244.x

31. Thomas A, Chin-Yee B, Mercuri M. Thirty years of teaching evidence-based medicine: have we been getting it all wrong? *Adv Health Sci Educ*. Published online September 24, 2021. doi:10.1007/s10459-021-10077-4

32. Mebius A. Corroborating evidence-based medicine. *J Eval Clin Pract*. 2014;20(6):915-920.

33. Tonelli. Integrating evidence into clinical practice: an alternative to evidence-based approaches: Integrating evidence into clinical practice. *J Eval Clin Pract*. 2006;12(3):248-256. doi:10.1111/j.1365-2753.2004.00551.x

34. Gupta M. Reconsidering rationality and ethics in the evidence-based medicine debate: a reply to commentators. *J Eval Clin Pract*. 2004;10(2):143-146. doi:10.1111/j.1365-2753.2004.00500.x

35. Lawson-Frost S. An epistemological problem for integration in EBM. *J Eval Clin Pract*. 2019;25(6):938-942. doi:10.1111/jep.13109

36. Mercuri M, Baigrie BS. What counts as evidence in an evidence-based world? *J Eval Clin Pract*. 2019;25(4):533-535. doi:10.1111/jep.13220

37. Miles A. On a Medicine of the Whole Person: Away from scientistic reductionism and towards the embrace of the complex in clinical practice. *J Eval Clin Pract*. 2009;15:941-949. doi:10.1111/j.1365-2753.2009.01354.x

38. Vere J, Gibson B. Evidence-based medicine as science. *J Eval Clin Pract*. 2019;25(6):997-1002. doi:10.1111/jep.13090

39. Lehane E, Leahy-Warren P, O'Riordan C, et al. Evidence-based practice education for healthcare professions: an expert view. *BMJ Evid-Based Med*. 2019;24(3):103-108. doi:10.1136/bmjebm-2018-111019

40. Roberts LW. Emerging Issues in Assessment in Medical Education: A Collection. *Acad Med*. 2021;96(2):159-160. doi:10.1097/ACM.0000000000003855

41. Shaneyfelt T, Baum KD, Bell D, et al. Instruments for Evaluating Education in Evidence-Based Practice: A Systematic Review. *JAMA*. 2006;296(9):1116-1127. doi:10.1001/jama.296.9.1116

42. Roberge-Dao J, Maggio L, Zaccagnini M, et al. The quality, methods and recommendations of systematic reviews on measures of evidence-based practice: an umbrella review. *JBI Evidence Synthesis*. 2022;20(4):1004-1073. doi:10.11124/jbies-21-00118

43. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657. doi:10.1007/s11136-011-9960-1

44. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing.*; 2014.

45. Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Chapter 10: Umbrella Reviews. In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis*. JBI; 2020. doi:10.46658/JBIMES-20-11

46. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336-341.

47. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101.

48. Braun V, Clarke V. *Successful Qualitative Research: A Practical Guide for Beginners*. sage; 2013.

49. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qual Res Sport Exerc Health*. 2019;11(4):589-597.

50. Campbell K, Orr E, Durepos P, et al. Reflexive Thematic Analysis for Applied Qualitative Health Research. *Qual Rep*. Published online June 20, 2021. doi:10.46743/2160-3715/2021.5010

51. Rowe WE. Positionality. In: Coghlan D, Brydon-Miller M, eds. *The SAGE Encyclopedia of Action Research*. Sage; 2014.

52. Lewis-Beck MS, Bryman A, Liao TF. Conceptualization, Operationalization, and Measurement. In: *The SAGE Encyclopedia of Social Science Research Methods*. Vol 1. SAGE Publications, Inc.; 2004:162-165. https://www-doi-org.proxy3.library.mcgill.ca/10.4135/9781412950589.n150

53. Tilson JK, Kaplan SL, Harris JL, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Educ*. 2011;11(1):1-10. doi:10.1186/1472-6920-11-78

54. Straus SE, Green ML, Bell DS, et al. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ*. 2004;329(7473):1029-1032.

55. Bandalos DL. *Measurement Theory and Applications for the Social Sciences*.; 2018. Accessed January 4, 2022. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1655848

56. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011.

57. Lewis CC, Mettert KD, Dorsey CN, et al. An updated protocol for a systematic review of implementation-related measures. *Syst Rev*. 2018;7(1):66. doi:10.1186/s13643-018-0728-3

58. Brownson RC, Colditz GA, Proctor EK. *Dissemination and Implementation Research in HealthTranslating Science to Practice*. Oxford University Press; 2012. doi:10.1093/acprof:oso/9780199751877.001.0001

59. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci*. 2014;9(1):118. doi:10.1186/s13012-014-0118-8

60. Nilsen P. Making sense of implementation theories, models, and frameworks. In: *Implementation Science 3.0*. Springer; 2015:53-79.

61. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *BMJ Evid-Based Med*. 2002;7(2):36-38.

62. Thomas A, Saroyan A, Dauphinee WD. Evidence-based practice: a review of theoretical assumptions and effectiveness of teaching and assessment interventions in health professions. *Adv Health Sci Educ*. 2011;16(2):253-276. doi:10.1007/s10459-010-9251-6

63. Satterfield JM, Spring B, Brownson RC, et al. Toward a Transdisciplinary Model of Evidence-Based Practice. *Milbank Q*. 2009;87(2):368-390. doi:10.1111/j.1468-0009.2009.00561.x

64. Kelly MP. The need for a rationalist turn in evidence-based medicine. *J Eval Clin Pract*. 2018;24(5):1158-1165. doi:10.1111/jep.12974

65. Tardif J. L'évaluation des compétences: documenter le parcours de développement. Published online 2006.

66. Saunders H, Gallagher-Ford L, Kvist T, Vehviläinen-Julkunen K. Practicing Healthcare Professionals' Evidence-Based Practice Competencies: An Overview of Systematic Reviews. *Worldviews Evid Based Nurs*. 2019;16(3):176-185. doi:https://doi.org/10.1111/wvn.12363

67. Simons MR, Zurynski Y, Cullis J, Morgan MK, Davidson AS. Does evidence-based medicine training improve doctors' knowledge, practice and patient outcomes? A systematic review of the evidence. *Med Teach*. 2019;41(5):532-538. doi:10.1080/0142159X.2018.1503646

68. Young T, Rohwer A, Volmink J, Clarke M. What Are the Effects of Teaching Evidence-Based Health Care (EBHC)? Overview of Systematic Reviews. *PLOS ONE*. 2014;9(1):e86706. doi:10.1371/journal.pone.0086706

69. Hatala R, Guyatt G. Evaluating the Teaching of Evidence-Based Medicine. *JAMA*. 2002;288(9):1110-1112. doi:10.1001/jama.288.9.1110

70. Savage CL, Kub JE, Groves SL. *Public Health Science and Nursing Practice: Caring for Populations*. F.A. Davis Company.; 2016.

71. Flake JK, Fried EI. Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Adv Methods Pract Psychol Sci*. 2020;3(4):456-465. doi:10.1177/2515245920952393

72. Belita E, Squires JE, Yost J, Ganann R, Burnett T, Dobbins M. Measures of evidence-informed decision-making competence attributes: a psychometric systematic review. *BMC Nurs*. 2020;19(1):44. doi:10.1186/s12912-020-00436-8

73. Newton PE, Shaw SD. Standards for talking and thinking about validity. *Psychol Methods*. 2013;18(3):301.

74. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ*. 2017;22(4):853-867. doi:10.1007/s10459-016-9716-3

75. Shankar S, Marshall SK, Zumbo BD. A Systematic Review of Validation Practices for the Goal Attainment Scaling Measure. *J Psychoeduc Assess*. 2020;38(2):236-255. doi:10.1177/0734282919840948

76. Association AER, Association AP, Education NC on M in, Educational JC on S for, Testing (US) P. *Standards for Educational and Psychological Testing*. Amer Educational Research Assn; 1999.

77. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3

78. Shaw S. Tracing the evolution of validity in educational measurement: past issues and contemporary challenges. 2011;(11):6.

79. Kane MT. Current concerns in validity theory. *J Educ Meas*. 2001;38(4):319-342.

80. Gorin JS. Reconsidering issues in validity theory. *Educ Res*. 2007;36(8):456-462.

81. Baldwin W. Information No One Else Knows: The Value of Self-Report. In: Stone AA, Bachrach CA, Jobe JB, Kurtzman HS, Cain VS, eds. *The Science of Self-Report: Implications for Research and Practice*. Psychology Press; 1999:3-7.

82. Bandura A, Wessels S. *Self-Efficacy*. Vol 4. na; 1994.

83. Thomas A, Kuper A, Chin-Yee B, Park M. What is “shared” in shared decision-making? Philosophical perspectives, epistemic justice, and implications for health professions education. *J Eval Clin Pract*. 2020;1(n/a). doi:10.1111/jep.13370

84. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Rethinking bias and truth in evidence-based health care. *J Eval Clin Pract*. 2018;24(5):930-938. doi:10.1111/jep.13010

85. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-1109. doi:10.1080/0142159X.2018.1500016

86. Colbert-Getz JM, Shea JA. Three key issues for determining competence in a system of assessment. *Med Teach*. 2021;43(7):853-855.

87. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? *J Rehabil Med*. 2012;44(2):97-98. doi:10.2340/16501977-0938

Acknowledgements

This study partially contributes towards the doctoral research of the first author (JRD). JRD acknowledges the generous support of a doctoral fellowship from McGill University (the Cameron-Herbert A. Davis Fellowship), the Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal (CRIR) and from the Richard and Edith Strauss Foundation.

This study was funded by the Canadian Institutes of Health Research (grant number 148544). The funding body did not influence the study results and was not involved in the study in any way.

Conflict of interest statement for all authors

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: JRD and AT had financial support from the Canadian Institutes of Health Research for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the U.S. Government.

Chapter 4: The Integration of Manuscripts 1, 2 and 3

Research objectives of manuscripts 1, 2 and 3

The global aim of the first study was to identify and describe the methods, results, recommendations, reported challenges and areas for future practice across systematic reviews (SRs) on evidence-based practice (EBP) measures in healthcare. The results from this study were reported in two manuscripts (Manuscripts 1 and 2).

Manuscript 1:

The first manuscript aimed to: (1) estimate the quality of SRs on EBP measures across health care professions and (2) identify differences between SRs regarding (a) methods used to assess the adequacy of EBP measures and (b) recommended measures.

Manuscript 2:

The second manuscript aimed to identify and describe the main challenges, recommendations for practice and areas of future research in the measurement of EBP across the health professions as reported by SRs.

Manuscript 3:

The global aim of this study was to describe the prototype development of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific study objectives were to: (1) identify candidate items best reflecting the most salient EBP domains for OTs and PTs across Canada from existing measures; (2) estimate the extent to which the prototype PIRE-CDMI "behaves" coherently across characteristics of the sample; and (3) estimate the extent to which the prototype PIRE-CDMI provides information that is comparable to that from other EBP measures.

Integration of manuscripts 1, 2 and 3

Manuscripts 1 and *2* reported on the first study, the umbrella review, which included 10 SRs on EBP measures across the healthcare professions. Findings from this umbrella review reinforced the need for a robust, comprehensive, and relevant measure of EBP and informed our reflective process for the next phase of developing the index. Three noteworthy considerations from the first study informed the next study.

First, findings show important limitations with existing measures including (1) the limited psychometric testing and evidence on existing EBP measures; (2) the absence of recommended measures in rehabilitation that comprehensively span all EBP domains of knowledge, skills, attitudes, and behaviors; and (3) the absence of measures including domains related to the organizational context.

Second, the review generated recommendations for the development of future EBP measures which included (1) the need for future EBP measures to be multidimensional and comprehensive, considering that all domains are interlinked; (2) the need for expert consensus on the operationalization of EBP; (3) the need to emphasise feasibility and acceptability of future measures.

Third, the findings revealed a tension in how validity is conceptualized in SRs. Specifically, most authors followed standards for patient-reported outcome measures and viewed validity as a property of the measure, while only one group of authors (who published most recently) applied the *Standards of Educational and Psychological Testing*[1] (hereafter referred to as *The Standards*) where validity is seen as a property of inferences and interpretations. This discrepancy could be attributed to the fact that psychometric guidelines in the clinical sciences are developed for patient-reported outcomes and the measurement target population of EBP measures are clinicians (i.e., not patients).

Thus, in taking the first steps towards developing the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI), it was essential to respond to this call for robust measures that could capture the multifaceted and complex nature of EBP.

Considering the findings from the umbrella review which revealed inconsistencies in the application of validity theory to EBP measures, I employed a contemporary perspective of validity wherein validity is regarded as a unified construct. *The Standards,* which is recognized as the gold standard for psychometric studies in educational, psychosocial, and health research settings, [1–3], is used as a frame of reference for the measure development process and assessing the validity of interpretations of results[1].

The four phases of the measure development process, as stated in the latest version of *The Standards* (2014), include: (1) development and evaluation of the measure specifications; (2) development, tryout, and evaluation of the items; (3) assembly and evaluation of new measure forms; and (4) development of procedures and materials for administration and scoring[1(p75)]. As the measure development process can be iterative, these phases may be repeated in various orders. *The Standards* articulates five categories of validity evidence (content[b], response process[c], internal structure[d], relations to other variables[e], and consequences[f]). The five-category framework for validity is highly referenced as the most comprehensive conceptualization of validity[3,4]. From this perspective, validity is seen as a property of scores and score interpretations within a specific context of evaluation rather than a property of the measure itself [1,3,5,6]. Validity is also seen as an ongoing activity whereby evidence sources support or refute the validity hypothesis.

The development of the prototype index described in *Manuscript 3* extends the work of Al Zoubi et al.[7] who initiated the development of a multidimensional measure of EBP by establishing an 70-item item pool from 181 initial items and confirming six core EBP domains. The authors then reduced the 70 items to 55 items using Rasch analysis[7]. *Manuscript 3* describes the process used to further reduce the 55-item pool to only five items (one item per EBP domain, minus one domain) resulting in a brief prototype index. Justification for continuing the measure development process lay in the necessity to test the hypothesis that this short prototype index provided comparable information to other measures of EBP. Therefore, the study described in *Manuscript 3* also provides validity evidence for scores generated by the prototype index with the sample of OTs and PTs relating to content, internal structure, and relationships to other variables.

---

[b] *Content* refers to the relationship between the content of the items (e.g., themes, wording, format, scoring) and the construct under measure, which includes analyzing the adequacy and relevance of items with respect to the content domain and the proposed interpretation of results

[c] *Response process* is defined as the analysis between the construct of interest and the cognitive processes of respondents when completing the measure

[d] *Internal structure* is defined as the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based

[e] *Relations to other variables* is defined as the extent to which relationships with external variables are consistent with the construct underlying the proposed test score interpretations

[f] *Consequences* involves gathering evidence to evaluate the soundness of proposed interpretations for their intended uses and can extend to other consequences beyond the interpretation of results

The research that will be presented in *Manuscript 3* corresponds to phases *a* and *b* of the four phases of measure development reported in *The Standards*[1]. In phase *a* (development and evaluation of the measure specifications), the purpose of the measure, intended users, construct and content inclusion and length specification are identified. In alignment with phase *b*, (development, tryout, and evaluation of the items), items were selected to meet the measure specifications (e.g., the purpose of the measure and the construct of interest) and the quality of items were evaluated through pretesting.

References

1. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing*.; 2014.

2. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, USA; 2015.

3. Zumbo BD, Chan EKH, eds. *Validity and Validation in Social, Behavioral, and Health Sciences*. Vol 54. Springer International Publishing; 2014. doi:10.1007/978-3-319-07794-9

4. Beckman TJ, Cook DA, Mandrekar JN. What is the Validity Evidence for Assessments of Clinical Teaching? *J Gen Intern Med*. 2005;20(12):1159-1164. doi:10.1111/j.1525-1497.2005.0258.x

5. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837.

6. Messick S. Validity. In: *Educational Measurement. American Council on Education Series on Higher Education.* 3rd ed. Oryx Press; 1993.

7. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

Chapter 5: Manuscript 3

Identifying Candidate Items for a Prototype Index on Propensity to Integrate Research Evidence into Clinical Decision-Making in Rehabilitation

Jacqueline Roberge-Dao[1,2], Aliki Thomas[1,2,3], Annie Rochette[2,4], Keiko Shikako[1,2], Nancy Mayo[1,5]

[1] School of Physical and Occupational Therapy, McGill University, Montréal, Canada
[2] Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Canada
[3] The Institute of Health Sciences Education, McGill University, Montréal, Canada
[4] L'École de réadaptation de l'Université de Montréal, Montréal, Canada
[5]Center for Outcomes Research and Evaluation (CORE), The Research Institute of the McGill University Health Center (RI-MUHC), Montreal, QC, Canada

Abstract

Objectives: Current measures of evidence-based practice (EBP) in rehabilitation provide a fragmented interpretation of EBP by discrete domains which conflicts with the multidimensional nature of EBP. Various individual and organizational factors synergistically affect whether a clinician will engage with the EBP process. The global aim of this study is to describe the prototype development of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific objectives are to: (1) identify from existing measures candidate items best reflecting the most salient EBP domains for occupational (OTs) and physical (PTs) therapists; (2) estimate the extent to which the prototype "behaves" coherently across characteristics of the sample; and (3) estimate the extent to which the prototype provides information that is comparable to that from other EBP measures.

Methods: Using a dataset from a large national study of Canadian OTs and PTs (N=127) who responded to a survey containing 70 items (representing six EBP domains) used to measure EBP, one item per domain was selected by visual inspection of Rasch analysis threshold maps and expert consensus. A scoring algorithm was developed using preliminary weights based on the logit placements on the Rasch continuum. A single score was generated to examine the interpretability of the prototype index score across characteristics of the sample of OTs and PTs and compared to full EBP measures using generalized estimating equations.

Results: Five items were selected for inclusion in the prototype PIRE-CDMI representing the dimensions of *use of research evidence, self-efficacy, resources, attitudes,* and *activities*. Testing of the prototype index demonstrated that it behaves consistently regardless of age groups, gender, and setting, and provides comparable information to full EBP measures in a succinct way.

Conclusion: This study describes the prototype development process of a brief, multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into clinical decision-making and demonstrates the benefits of having an overall indicator of EBP as opposed to using various measures. This study provides a proof-of-concept measure and evidence that such a measure would be valuable.

Introduction

Evidence-based practice

Clinical decisions in rehabilitation can reflect the quality of rehabilitation services offered and directly impact patient outcomes[1]. Evidence-based practice (EBP), which has become synonymous with high-quality effective care, is commonly defined as an approach to clinical decision-making (CDM) which integrates research evidence, clinical expertise and patient preferences within the context of available resources[2–4]. As such, it is an expectation that rehabilitation clinicians including occupational (OTs) and physical therapists (PTs) engage in EBP [5–8].

In recent years, studies have shown that although rehabilitation professionals' attitudes towards EBP are generally quite favorable, the use of research evidence is still a challenge[9–11]. Several individual and organizational factors have been found to be associated with a rehabilitation clinician's use or non-use of research evidence[12–16]. Individual factors include limited self-efficacy in applying research evidence to practice[17,18], lack of familiarity with evidence based interventions[19] and difficulty forming new habits using novel interventions[14]. It has also been suggested that participation in formal (e.g., courses, conferences) and informal (e.g., reading the literature) learning activities may also have an impact on the likelihood of engaging in EBP[13,20–25]. Organizational factors can also either facilitate or hinder EBP[10,26–32]. In a 2021 systematic review and meta-analysis of 29 articles on barriers to EBP in physiotherapy across a range of practice settings, Paci and colleagues found that the principal barrier was lack of protected time for activities related to EBP such as reading the literature followed by the lack of access to resources such as research databases[33]. Similarly, a 2020 scoping review in occupational therapy highlighted lack of staff and EBP experts, increased costs associated with specific interventions, lack of time, logistical challenges, and inadequate equipment as being important organizational barriers to EBP[14]. These organizational barriers can be mitigated by allocating more time for clinicians to access and appraise research evidence and providing sufficient human (e.g., EBP expert) and material (e.g., equipment) resources[19].

The measurement of EBP

Given the call for EBP, robust measurement practices are essential to identify EBP engagement levels of clinicians and to identify the factors[g] related to EBP that should be improved or maintained. The measurement of EBP can serve to determine the effectiveness of interventions aimed at increasing research uptake and identify clinicians' needs for targeted allocation of resources to support EBP. Ultimately, precise measurement of EBP supports quality of care by informing researchers, clinicians, and managers on areas for improvement.

More than 200 EBP measures were identified in a recent umbrella review (chapters 3.1), highlighting a growing interest in this area of research[34]. In occupational and physical therapy, there exist at least 30 self-report questionnaires which assess various EBP domains[h36,37]. Some of these questionnaires cover a single EBP domain such as self-efficacy (*Evidence-based Practice Confidence Scale* by Salbach et al.[18]) or attitudes (*Questions on EBP questionnaire* by Stevenson et al.[38]) while others[28,39–51] combine multiple domains in various combinations such as the *Evidence-Based Practice Profile Questionnaire* (EBP^2) by McEvoy et al.[52] which covers five domains: Relevance, Terminology, Confidence, Practice, and Sympathy. Given the various individual and organizational factors which, together, influence whether a clinician will integrate research evidence into CDM, it has been suggested to include all the key EBP domains into a single measure to gain a more comprehensive interpretation of EBP[37,53]. Among the EBP measures recommended for use in rehabilitation derived from three systematic reviews[34,36,37,54], none comprehensively cover all domains associated with EBP, namely knowledge, attitudes, skills and behaviors. Furthermore, the authors of one of the reviews reported that none of the questionnaires measured the organizational factors that may influence the enactment of EBP[37]. Since the publication of the three systematic reviews in 2010[54], 2016[36] (occupational therapy) and 2014[37] (physical therapy), two new measures of EBP in rehabilitation have been published including the *Health Sciences Evidence Based Questionnaire*[39] and the measure by Al Zoubi et al. [55]. Unlike earlier questionnaires, these recent questionnaires are comprehensive and include items assessing the organizational-level factors known to influence EBP.

---

[g] In this study, *factor* is used to describe a personal or organizational influence on a clinician's propensity to integrate research evidence into clinical decision-making.

[h] *Domain* is used in the context of measurement to describe a conceptually defined part of a construct. *Domain* can also be used to represent a subscale of a measure[35].

An important challenge with the majority of EBP questionnaires identified in the systematic reviews lies in the inappropriate treatment of ordinal data. Most EBP questionnaires consist of Likert-type scales whereby respondents select textual responses rather than numbers on a scale. Often, numbers are attributed to the categories of the Likert-type scale and authors sum or average across items[56,57]. For instance, the EBP[2] by McEvoy et al.[52] includes 74 items, covering five domains, all of which used a five-point Likert scale (e.g., not at all true, not really true, possibly true, quite likely true, very true). In the analysis, the ordinal data are treated as interval-level data as the authors present mean scores, standard deviations and effect sizes[58]. Such an approach to analysis is flawed because the allocation of numbers is not indicative of any mathematically valid magnitude of difference between categories and performing arithmetic operations on ordinal data can produce misleading conclusions[57,59].

The research by Al Zoubi and colleagues[55] addressed this challenge by using Rasch analysis, a modern measurement approach, to determine the mathematical properties of six unidimensional scales of EBP. Items for each EBP domain were calibrated onto the same linear scale which allowed for a mathematically valid approach to summing scores. In doing so, the measure[i] by Al Zoubi et al.[55] includes all core EBP domains (knowledge, use of research evidence, self-efficacy, resources, attitudes, and activities related to EBP) and generates total scores for each domain given the interval-level nature of the scales.

Despite these important developments in the measurement of EBP (i.e., the inclusion of individual *and* organizational EBP domains and the use of modern measurement theory), existing EBP measures can be considered *profiles*[35] whereby results are summarized by sub-scale or EBP domain. This fragmented interpretation of EBP by individual domains conflicts with the multidimensional nature of EBP whereby individual and organizational factors are likely to synergistically affect whether a clinician will engage with the EBP process. The failure to combine EBP domains for an overall indicator of EBP is one of the major shortcomings with existing EBP measures. Considering that there is always a need for parsimony in measurement, a concise yet comprehensive measure able to generate an overall indicator of EBP in rehabilitation is needed.

---

[i] The term *measure* is used intentionally because the set of included items have been shown to form a unidimensional linear continuum and have interval-level properties[35].

Further, as mentioned in chapter 1, purporting to measure EBP directly becomes problematic due to a more modern conceptualization of EBP which advocates for including more than just research evidence. Indeed, other components such as professional expertise, expert opinion, and patient preferences and values are increasingly recognized as contributing to the EBP process [60–65]. Most of the literature claiming to measure EBP does not account for other components encompassed in the EBP process outside research evidence. Unless all components of EBP are accurately captured, it is unjustified to claim *EBP* questionnaires or *EBP* measures.

Objectives

The global aim of this study is to describe the prototype development of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific study objectives are to (1) identify candidate items best reflecting the most salient EBP domains for OTs and PTs across Canada from existing measures; (2) estimate the extent to which the prototype PIRE-CDMI "behaves" coherently across characteristics of the sample; and (3) estimate the extent to which the prototype PIRE-CDMI provides information that is comparable to that from other EBP measures. As it will not have been subjected to qualitative revision, weight elicitation methods or field testing, this descriptive system will be referred to as the prototype index or the P-PIRE-CDMI. This study is designed to produce a proof-of-concept measure and to provide evidence to justify the need to develop a deployable measure.

Methods

Conceptual framework

EBP is a latent construct, a variable that cannot be directly observed but can be estimated and inferred by related variables[35,66]. Measuring a clinician's *propensity* to integrate research evidence into CDM may be one approach to estimating and inferring EBP. In a conceptual model where *EBP* is the exposure and *improved patient health* is the outcome, propensity variables predict the exposure (EBP)[67]. In this dissertation, *propensity* is described as a rehabilitation clinician's conditional probability or likelihood of integrating research evidence into CDM given specific individual and organizational determinants. Propensity to integrate research evidence into CDM (an important element of EBP) can be characterized through set features that are well-established in the literature[68]. As described in chapter 1, there exists a well-articulated, accepted body of knowledge with respect to the influence of certain individual and organizational factors on EBP and to the measurement of those factors. The term *propensity* is defined by the Cambridge Academic Content Dictionary as "a tendency to behave in a particular way". Using propensity as a proxy for behavior acknowledges the dynamic and uncertain nature of EBP and provides a practically relevant approach to measuring a complex construct.

In applied statistics, propensity modeling is a set of techniques which attempts to estimate the likelihood of subjects performing a behavior by accounting for variables that affect such behavior[69]. In the context of EBP, measuring propensity requires a clear definition of the core variables that influence a clinician's integration of research evidence into CDM. This study will use the previously identified six core EBP domains in rehabilitation as the propensity variables (knowledge, use of research evidence, self-efficacy, resources, attitudes, and activities related to EBP)[55]. The underlying measurement model for the construct of interest (propensity to integrate research evidence into CDM) is a formative model, meaning that dimensions form the construct in an additive way (i.e., individual and organizational factors cause or collectively produce the latent phenomenon) as opposed to reflecting the construct as in a reflective model[55,70]. As such, the most appropriate approach in generating a total score for formative constructs is a weighted sum of dimensions, assuming that the dimensions are independent (i.e., have little correlation between items).

Modern measurement approaches are increasingly being used to develop and examine measures in healthcare as these approaches analyze the mathematical properties of items and allow for item reduction [71–76]. Rasch Measurement Theory (RMT) is a modern measurement approach which allows for the mapping of unidimensional constructs onto a single, linear scale from the easiest to the most difficult item using a logit transformation[77–79]. RMT estimates the spread of respondents on a continuum based on their ability levels. Rasch analysis can assess item fit to the Rasch model, whether items span the entire construct, the extent to which items have response choices that are appropriately ordered and whether items perform differently between groups of people[76,80,81]. Rasch analysis results allow researchers to reduce the number of items and rewrite or exclude problematic items (i.e., items prone to bias). This allows for a more rapid and less burdensome test experience. Furthermore, calibrating items onto the same linear scale allows for a mathematically valid approach to summing scores. Finally, using the Rasch approach, responses to single items can be selected to reflect the latent construct rather than relying on the entire scale[79].

Data source

This study consists of a secondary analysis of an existing dataset that that sought to measure use of research evidence (as the primary outcome) longitudinally in newly graduated OTs and PTs from 12 rehabilitation programs in Canada[11,55]. This sample was used because they completed the multidimensional measure by Al Zoubi et al.[55] at four time points (each time point was one year from the next) which served as a foundation for developing the prototype index.

The detailed process for developing the survey is described elsewhere[55]. In sum, the research team assembled 181 items from the five following commonly used measures of EBP selected based off a literature review and expert consensus: the *Evidence-Based Practice Profile Questionnaire* (also called the EBPQ)[82], the *Evidence-Based Practice Profile Questionnaire-2* (also called the EBPQ2)[52], the *Evidence-Based Practice Attitude Scale* (also called the EBPAS)[83], the *Evidence-Based Practice Confidence Scale* (also called the EPIC)[84] and the *Alberta Context Tool* (also called the ACT)[85].

These 181 items were reduced to 70 items through a nominal group technique and a modified Delphi process. The 70 items were individually classified by the research team into six EBP core domains which were then labelled and categorized based on their underlying conceptual model (reflective or formative). Four constructs (*knowledge, self-efficacy, resources,* and *attitudes*) were found to be reflective, meaning that the items are influenced by or are caused by the latent construct[35]. Given that summing items of a reflective construct may be arbitrary, this classification signaled the need to examine the mathematical properties of these reflective constructs to develop a robust and accurate scoring system. Two constructs (*use of research evidence* and *activities related to EBP*) were found to be formative, meaning that the items form or make up the latent construct. In such cases, summing across items is acceptable because items jointly combine to produce the latent phenomenon[35,70].

At baseline (T0), new graduates responded to this 70-item online survey. Rasch analysis was conducted on the responses to the four reflective constructs which allowed for item reduction by fixing or omitting certain items that did not fit the Rasch model. For example, Rasch analysis ruled out differential item functioning by language (English and French). The three subsequent surveys (each a year apart) were reduced to 55 items stemming from six EBP domains - *Knowledge* (8 items, 0-29 scale), *Use of research evidence* (9 items, 5-9), *Self-efficacy* (8 items, 0-22), *Resources* (13 items. 0-39), *Attitudes* (10 items, 0-32), and *Activities related to EBP* (7 items, 1-5). This item reduction process ensured that scores from these unidimensional measures have a conceptually defensible meaning that supports their interpretation and use[86].

Sample characteristics

The sample used for analysis in this study consisted of newly graduated OTs and PTs from 12 rehabilitation programs in Canada that (1) were working at the time of survey completion and (2) completed all items of the survey at baseline, T0 (n=127) and four years later, T3 (n=37). Table 1 summarizes the characteristics of the sample (mean age: 27; SD: 2.7-3.9, predominantly women).

Table 1: Descriptive demographics of the sample at baseline and three years later

Sample characteristics	Mean (SD) or n (%)	
	Used for analysis at T0	**Used for analysis at T3**
	(N=127)	(N=37)
Age (years) at baseline	27 (2.7)	27 (3.9)
Women/Men/Did not specify	101/24/2 (80/19/2)	32/4/1 (87/11/3)
English/French	82/45 (65/35)	19/18 (51/49)
Physical therapist/Occupational therapist	74/53 (58/42)	15/22 (41/60)
Work setting		
General hospital - acute care	11 (9)	6 (16)
Long term care or rehabilitation center	7 (6)	1 (3)
Community agency, primary health care, home visiting agency, consulting firm	26 (21)	10 (27)
Private practice	80 (63)	18 (49)
Missing or not applicable	3 (2)	2 (5)
Continuous variables are presented as means and standard deviation (SD) while categorical variables are presented as frequencies (%).		

Data analysis and procedure

Both Classical Test Theory[87] and RMT are used in this study. RMT serves to select the best performing item per EBP domain and Classical Test Theory is used to comment on the performance of the prototype index using the existing dataset.

A priori objectives for developing the prototype

The authors set out to develop a concise measure which would reduce administration time and response burden for clinicians while covering the multiple domains that form the latent construct: propensity to integrate research evidence into CDM. It was expected that each of the six EBP domains would be represented by the minimum number of items and response options for ease of completion of the index while still maintaining a discriminatory ability in item responses. Specifically, the aim was to select one item per EBP domain to align with the methodology associated with the development of preference-based measures for patients[76,78,79,88–90].

Item selection

The best performing item for each reflective construct was selected based on the Rasch analysis threshold maps by Al Zoubi et al.[55] using the procedure described elsewhere[76,91] (see Appendix I for thresholds maps). Threshold maps are visual representations of the ordinal responses transformed onto a linear scale using a logit transformation. This logit scale depicts the person-item difficulty of a construct where negative logit values indicate people that have more difficulty with the item while positive logit values indicate less to no difficulty with the item. The visual representation of threshold locations across the latent trait allows the elimination of items with disordered thresholds. Disordered items correspond to items where respondents cannot distinguish between item levels. Two team members (JRD, NM) reviewed the threshold maps and selected an item per EBP domain which had interval-like response options, spanned the continuum, and centered on logit zero. Item levels could be collapsed or combined (reducing the number of item levels) if they were closer together compared to neighboring levels to obtain interval-like response options.

The remaining research team members (AT, KS, AR) were part of the expert panel that helped to validate the choices for each of the four reflective constructs. For each of the two formative constructs, one item was selected through discussion and consensus[92]; this panel considered the theoretical relevance and representativeness, clarity, length, directionality, distribution of responses (e.g., absence of floor or ceiling effects) and potential biases of items[93]. In this context of developing a multidimensional index and for purposes of clarity, the selected item best representing the EBP domain is hereafter called a *dimension*[78].

Development of a scoring algorithm

A scoring algorithm was developed to obtain a score for each participant and to test the performance of the prototype index. The scaled score method was used to assign preliminary weights to each EBP dimension-level[91]. The logit placements on the Rasch linear continuum were used to develop weights that represented the interval spacing for each of the response options. The original logit ability estimates were recalibrated so that the lowest logit score represented zero. To obtain a total score for a participant, the points associated with their response option for each of the five dimensions were summed in a simple additive formula. These weights are not to be confused with preference weights or utilities as participant preferences have not been elicited at this stage. The preliminary weights are based on the actual responses from T0 and provide an indication of the actual performance of respondents along the continuum of item difficulty. Total scores were transformed to a 0-100 scale to ease interpretation and comparisons with the initial EBP measures which were also transformed onto 100. On the prototype index, zero represents neutral propensity to integrate research evidence into CDM (because a clinician cannot have no propensity) and 100 represents the highest propensity. A three-point classification level was attributed to each response option to denote best (A), middle (B), and worst (C) to allow for the generation and interpretation of EBP profiles based on P-PIRE-CDMI responses. Profiles were categorized hierarchically based on the number of middle and worst response options. For instance, if a respondent scored ABAAA, then the profile would be classified as having one B.

Polychoric correlation coefficients were calculated to ensure that dimensions were not highly correlated with each other. We hypothesized that the correlation coefficients would be low to moderate (no greater than 0.6); otherwise, it may signal a need to eliminate one of the items to increase structural independence.

Analysis of the prototype across characteristics of the sample

Descriptive statistics were used to determine the distributional parameters for the P-PIRE-CDMI and the EBP measures. For this paper, we assume that a minimal important difference is about half a SD of raw scores[94]. The standardized response mean (SRM) was calculated by dividing the mean difference in score from T3 to T0 by the standard deviation (SD) of the mean difference. The SRM is a measure of effect size that uses SD units allowing for comparison between measures[95]. Cohen's criteria were used to interpret the magnitude of the SRM, where 0.2 is small, 0.5 is moderate, and 0.8 is large[96]. Paired *t*-tests were also conducted to estimate the average difference at both timepoints for each measure.

Simple linear regression (SLR) was conducted to estimate the effect of single characteristics on the prototype score, at both time points. These characteristics were treated categorically: age (younger, middle, and older), gender (women, men, did not specify), profession (OT, PT) and setting (private practice, acute general hospital, long term care/rehabilitation center, and community/home visiting agency); age was also treated continuously.

Generalized estimating equation (GEE) analyses were conducted (exchangeable correlation structure) to determine if there were any important interactions which would signal a potential source of variance over time by characteristic. The model was: P-PIRE-CDMI score = [characteristic (age centered, age group, gender, profession, or setting) + time + characteristic*time]. GEE incorporates a within-subject correlation structure and accounts for correlations in scores within the same individual at both timepoints (i.e., in this model, we fit *time* as the within-subject variable). The main effect of time was not interpreted in the model given that our analysis only included two timepoints which is not a reliable estimate of change, and this was not the aim of this study; we were specifically interested in the potential interaction of *characteristic*time*.

Analysis of the prototype compared to EBP measures

Spearman Rho was used to correlate scores from the P-PIRE-CDMI and the five unidimensional EBP measures. It was hypothesized that measures would be moderately correlated (between 0.3 and 0.6) because the P-PIRE-CDMI is multidimensional, encompasses items included in the EBP measures and was completed by the same sample.

GEE analyses were conducted to determine the extent to which average scores of the EBP measures differed from the P-PIRE-CDMI. The model was: Score = [measure] (whereby the measures included the P-PIRE-CDMI, as referent, and the five unidimensional EBP measures). GEE was used to control for the correlation from multiple measurements on the same individual (i.e., in this model, we fit *measure* as the within-subject variable).

To determine the extent to which the P-PIRE-CDMI provides information that is comparable to EBP measures across characteristics of the sample (age, gender, profession, setting or university), the data were analyzed two ways: parametrically and non-parametrically. This dual analysis was conducted because of the ordinal nature of two EBP measures which violated the assumption of normality. First, GEE was conducted (parametric); the model was: Score = [measure + characteristic + measure*characteristic]. This analysis informed on whether there was a characteristic effect across all measures and whether the effect of measure depended on characteristic. All variables were treated categorically except age which was also treated continuously and was centered for analysis. Second, an analysis of ranks using chi-square was conducted to determine if the measures ranked characteristic groups in different ways (non-parametric).

To determine if the P-PIRE-CDMI behaves in a similar way to the other measures over time, the SRM was calculated, and paired t-tests were conducted. It is important to acknowledge that the SRM and *t*-test underestimate the variability due to the clustering of multiple measurements at the individual level at two time points. To account for clustering at the individual level across both timepoints, GEE was conducted to determine the presence of an interaction effect between measure and time; the model was: Score = [measure + time + measure*time]. Logistic regression was used to estimate whether age, gender, profession, setting, university, and language were associated with dropout rates from T0 to T3. SPSS Statistics Version 28 was used for all statistical analyses in this study.

Results

The prototype index

Five items were selected for inclusion in the P-PIRE-CDMI. The items, response options, threshold logit estimates, score transformations and classification labeling are presented in Table 2. Figure 1 illustrates the frequency distribution of P-PIRE-CDMI profiles based on the hierarchical classification system: no individuals were at the theoretical best or theoretical worst. The polychoric correlations between included items are presented in Appendix II and ranged from 0.34 to 0.57. Our estimate of a minimal important difference is 6.5, which represents half a SD of average P-PIRE-CDMI scores at baseline (12.9).

The expert panel agreed to omit the *knowledge* domain for three reasons. First, the items in the initial measure only represented knowledge about statistical concepts which is not the only type of knowledge invoked when integrating research evidence into CDM (other types of knowledge include knowledge on the five-step EBP process, knowledge of best practice evidence[63,97]). Second, knowledge is best assessed through cognitive testing not self-report (i.e., an individual could score high and believe that they understand the statistical term, when in fact they do not actually understand it[98(p2)]). Indeed, a ceiling effect was found for the full *knowledge* measure wherein 84% of new graduates rated themselves very knowledgeable about statistical concepts[55]. Third, previous findings using this dataset suggest that *knowledge* was not related to *self-reported use of research evidence* nor *activities related to EBP* upon entry into practice[11]. All the above converge to suggest that knowledge related to EBP cannot be measured using one item.

Appendix III presents the distribution of scores of EBP measures (including the P-PIRE-CDMI and the five unidimensional EBP measures) across characteristics of the sample at T0 (N=127). It also reports the mean scores at T0 and T3 (n=37), the average difference in means over time, the standardized response mean and *t*-test results. No association was found between age, gender, profession, setting, university, and language with dropouts from T0 to T3.

Figure 1. Distribution of the prototype index profiles at baseline

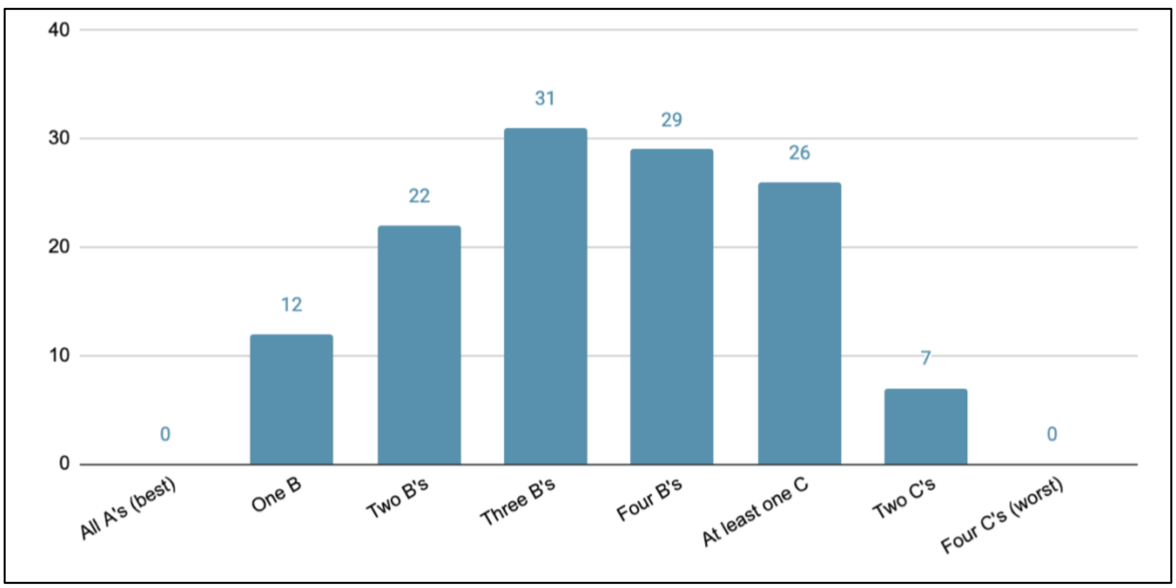


Table 2. Candidate items and response options, score transformations and classification system

Dimension	Question	Selected item	Response options	Threshold logit scores	Threshold logit scores recalibrated	Recalibrated logit on 100	Scoring on 10	Classification label A-B-C
Use of research evidence	How often have you done each of the following activities in the past month?	Decide on an appropriate course of action based on integrating the research evidence, clinical judgment and patient or client preferences?	0 (Never)	Formative model– no Rasch analysis conducted		0	0	C
			1 (One time or more)			50	5	A
Self-efficacy	Please indicate how confident you are in your current level of ability by choosing the corresponding number on the following rating scale.	Determine if the evidence from the research literature applies to your patient	0 (0-25%)	NA	NA	0	0	C
			1 (25-50%)	-2.834	3.2	26.4	2.6	B
			2 (50-75%)	-0.136	5.9	48.8	4.9	B
			3 (75-100%)	2.466	8.5	70.5	7	A
Resources	Please indicate your level of agreement with the following statements.	My organization supports best practice	0 (Strongly disagree /disagree)	NA	NA	0	0	C
			1 (Neutral)	-2.028	4.0	33.1	3.3	B
			2 (Agree)	-0.274	5.7	47.7	4.8	B
			3 (Strongly agree)	1.845	7.8	65.3	6.5	A
Attitudes	Please indicate your level of agreement with the following statements.	I am willing to use new and different types of clinical interventions (e.g., assessment, treatment) developed by researchers to help my patients/clients	0 (Strongly disagree)	NA	bottom	0	0	C
			1 (Disagree)	-5.835	0.2	1.4	0	C
			2 (Neutral)	-0.282	5.7	47.6	4.8	B
			3 (Agree)	-0.114	5.9	49.0	4.8	B
			4 (Strongly agree)	1.74	7.7	64.5	6.5	A
Activities related to EBP	In the past month, how often have you…?	Made time to read research	0 (Never)	Formative model – no Rasch analysis conducted		0	0	C
			1 (Monthly or less)			25	2.5	B
			2 (Bi-weekly)			50	5	B
			3 (Weekly)			75	7.5	B
			4 (Daily)			100	10	A

The prototype index across characteristics of the sample

Figures 2 to 5 illustrate the distribution of P-PIRE-CDMI scores across age group, gender, profession, and setting, respectively, at both timepoints. Appendix IV presents the SLR estimates for the modeling of P-PIRE-CDMI scores as a function of characteristic, at both time points individually. Appendix V reports the GEE estimates for the modeling of P-PIRE-CDMI scores as a function of characteristic, time, and the interaction.

There was no effect of gender or setting on P-PIRE-CDMI scores at both timepoints as suggested by SLR, and no important interaction of these individual characteristics with time in the GEE model.

An important effect was found for profession at baseline whereby OTs scored, on average, 7.1 points lower on the P-PIRE-CDMI (95% CI: -11.5 to -2.6) compared to PTs. Three years later, there was no longer an effect of profession on P-PIRE-CDMI scores (β= -0.8, 95% CI: -9 to 7.4). GEE signaled an important interaction term of profession*time.

SLR was suggestive of an age effect on P-PIRE-CDMI at baseline when treated continuously (β = -0.8, CI: -0.7 to 0.0). GEE was suggestive of an age effect when treated categorically (older age group: β = -11, CI: -19.6 to -2.4) and continuously (β = -1.6, CI: -2.2 to -0.9). After excluding three outliers from the sample of 127 participants at T0 and 37 participants at T3 (age at baseline: 36, 36 and 43), SLR no longer detected an effect of age at either time point, but GEE still found an effect of age when treated continuously (β= -1.7, CI: -3.4 to 0.0).

Figure 2. Distribution of prototype index scores across age groups at both timepoints

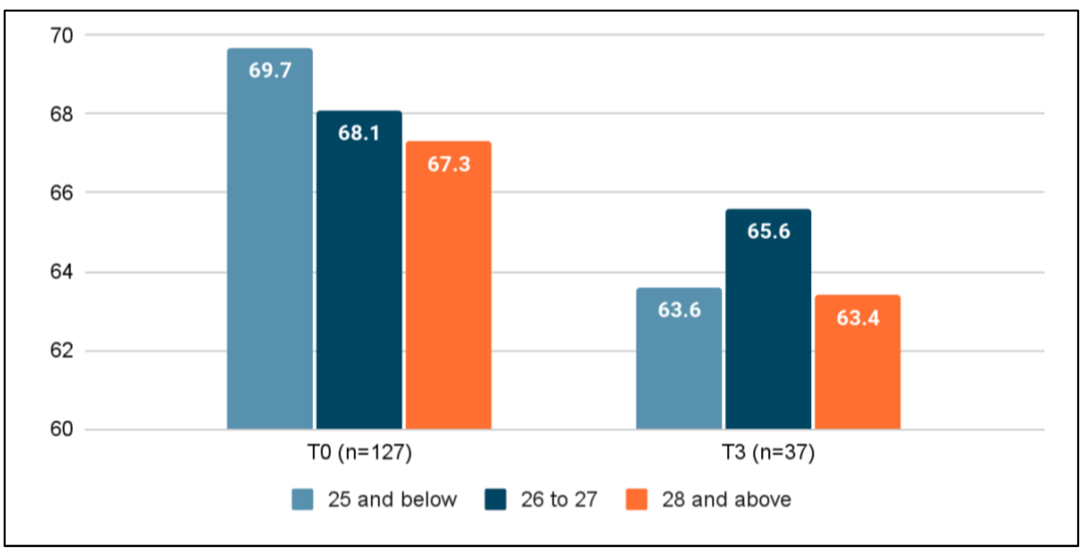


Figure 3. Distribution of prototype index scores across gender at both timepoints

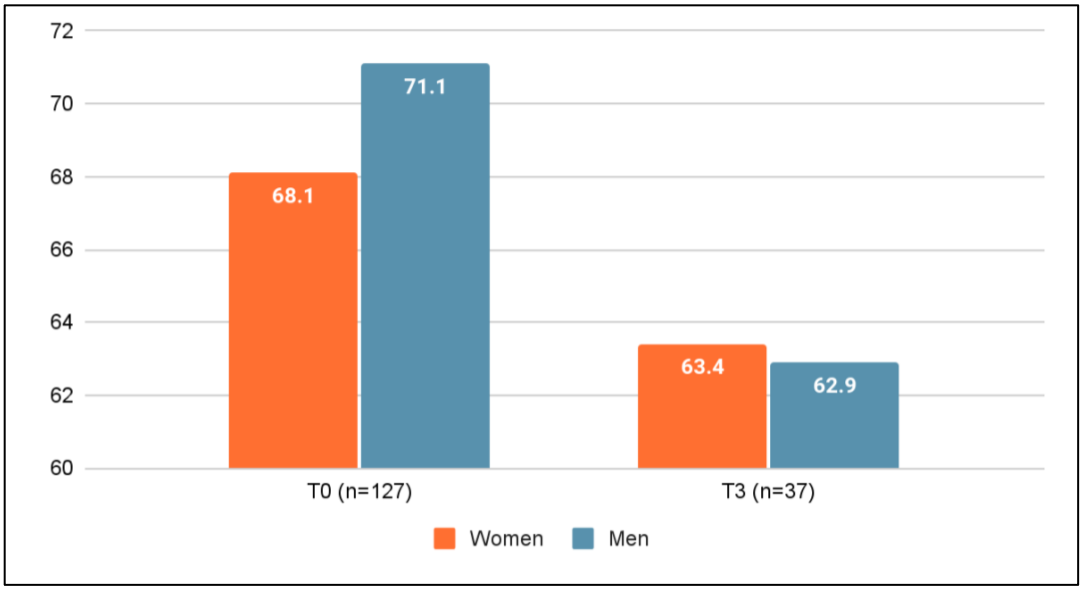


Figure 4. Distribution of prototype index scores across profession at both timepoints

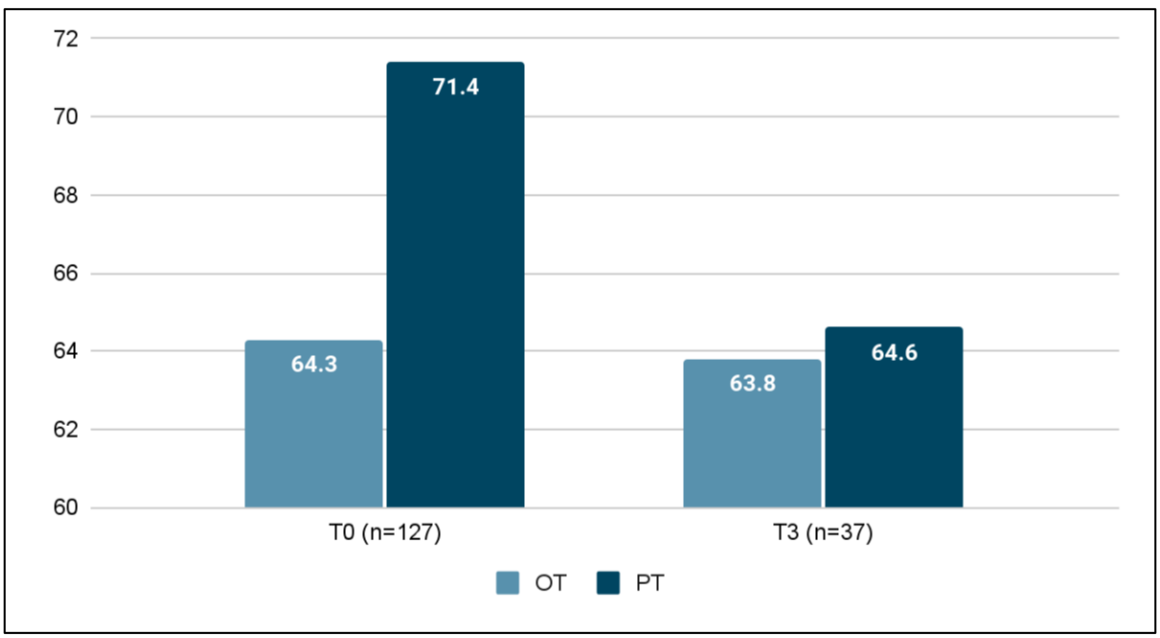
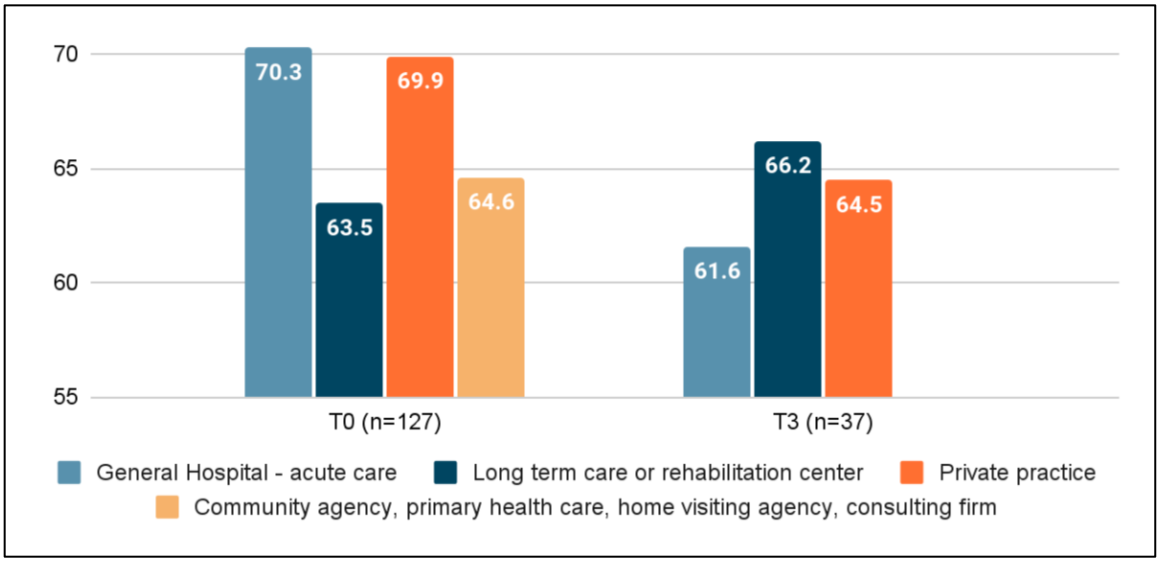


Figure 5. Distribution of prototype index scores across setting at both timepoints



The prototype index compared to other EBP measures

The P-PIRE-CDMI correlated moderately with the *self-efficacy* (r=0.36, [95% CI: 0.19 to 0.51]), *resources* (r=0.37, [95% CI: 0.21 to 0.52]), *use of research evidence* (r=0.46, [95% CI: 0.30 to 0.59]), *attitudes* (r=0.54, [95% CI: 0.40 to 0.65]) and *activities* (r=0.62, [95% CI: 0.49 to 0.72]) measures. Table 3 reports the GEE estimates illustrating the difference in average scores between measures at baseline.

Table 3. GEE estimates for scores as a function of EBP measure at baseline (N=127)

	Estimate (ß)	**SE**	**95% CI**		**t-Statistic [a] (ß /SE)**
			Lower	**Upper**	
P-PIRE-CDMI (referent)	68.4	1.1	66.2	70.7	
Activities [b]	-55.7	1.2	-57.9	-53.4	-48.1*
Resources	-3.6	1.4	-6.2	-0.9	-2.6*
Use of research evidence	11.9	1.4	9.1	14.7	8.4*
Self-Efficacy	-5.9	1.5	-9.0	-2.9	-3.8*
Attitudes	0.9	1.0	-1.1	3.0	0.9

GEE model: Score=measure
Dependent Variable: Score on each measure (out of 100) at baseline

GEE, generalized estimating equations; ß , beta; SE, standard error; CI, confidence interval; P-PIRE-CDMI, prototype of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index*
*There was an effect of the measure
[a] The t-statistic is a metric of effect size.
[b] Lower beta due to the transformation of the ordinal scale onto a 0-100 scale.

Figures 6 to 9 portray the distribution of the P-PIRE-CDMI and five EBP measures by age group, gender, profession and setting, respectively, at baseline. Appendix VI outlines the GEE estimates of measure; measure and characteristic; and measure and time on scores.

There was an important effect of age (β: -0.85, 95% CI: -1.66 to -0.04) across measures and an important interaction effect of age with the *attitudes* (β: 1.1, 95% CI: 0.5 to 1.7) and *activities* (β: -0.87, 95% CI: 0.19 to 1.5) measures. No important difference in ranks of measures by age group was found ($\chi 2$ (df:2, .05) =4.06, p=0.13). Our findings also reveal an important interaction effect of gender with the *self-efficacy* (β: 8.7, 95% CI: 3.9 to 1.1) measure. No important difference in ranks of measures by gender was found (χ2 (df:2, .05) =3.25, p=0.2). Results suggests an important overall profession effect across measures (β: -7.1, 95% CI: -11.5 to -2.6). No difference in ranks of measures by profession was found (χ2 (df:2, .05) =0, p=1). There was an important interaction effect of setting with the *resources* (β: 8.1, 95% CI: 0.6 to 15.6 and β: 14.6, 95% CI: 5.8 to 23.3) and *activities* (β: 6.5, 95% CI: 0.1 to 13.0 and β: 5.6, 95% CI: -1.0 to 12.2) measures. An important difference in ranks of measures by setting was found (χ2 (df:4, .05) =16.3, p=.003). Similarly, there was an important interaction effect of universities with the *self-efficacy* (β: 9.8, 95% CI: 0.4 to 19.2; β: -11.5, 95% CI: -21.3 to -1.8; β: -15.9, 95% CI: -27.8 to -3.9; and β: 11.5, 95% CI: 3.6 to 19.5) and *resources* (β: 13.3, 95% CI: 3.4 to 23.2) measures and a very important heterogeneity in the ranking of universities by measure (χ2 (df:9, .05) =55.58, p=9.4E-9). Finally, our results did not find an important interaction effect of measure with time meaning that there was no difference in the way measures performed from baseline to T3.

Figure 6. Distribution of scores from the prototype and five EBP measures by age group at baseline

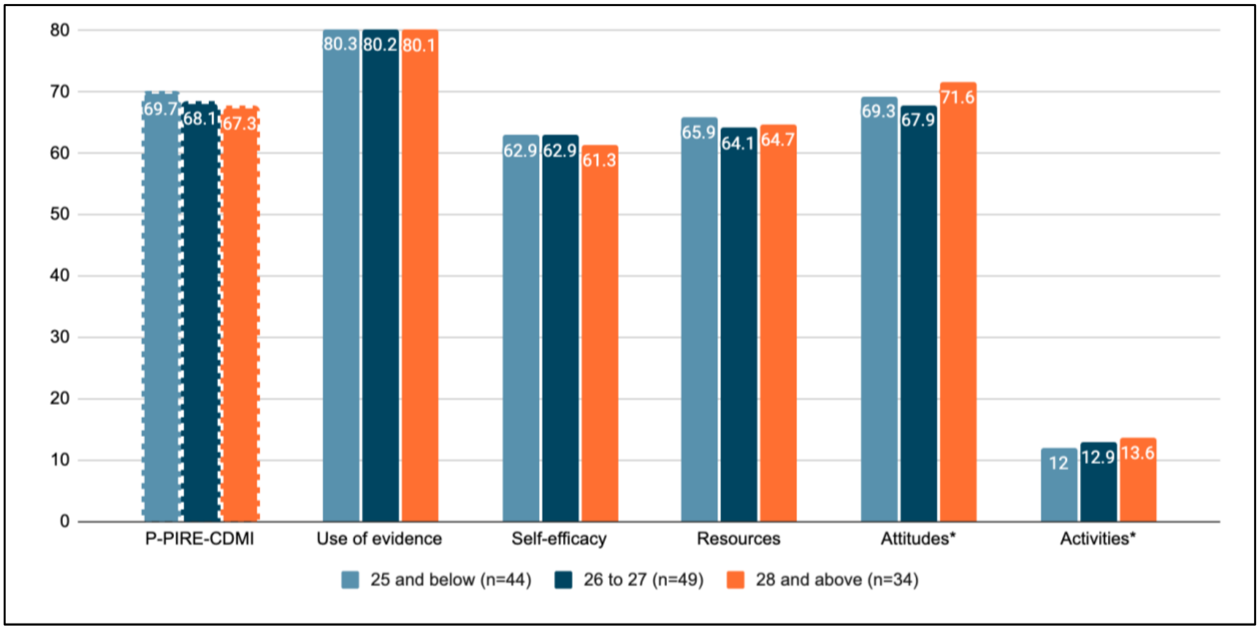


Figure 7. Distribution of scores from the prototype and five EBP measures by gender at baseline

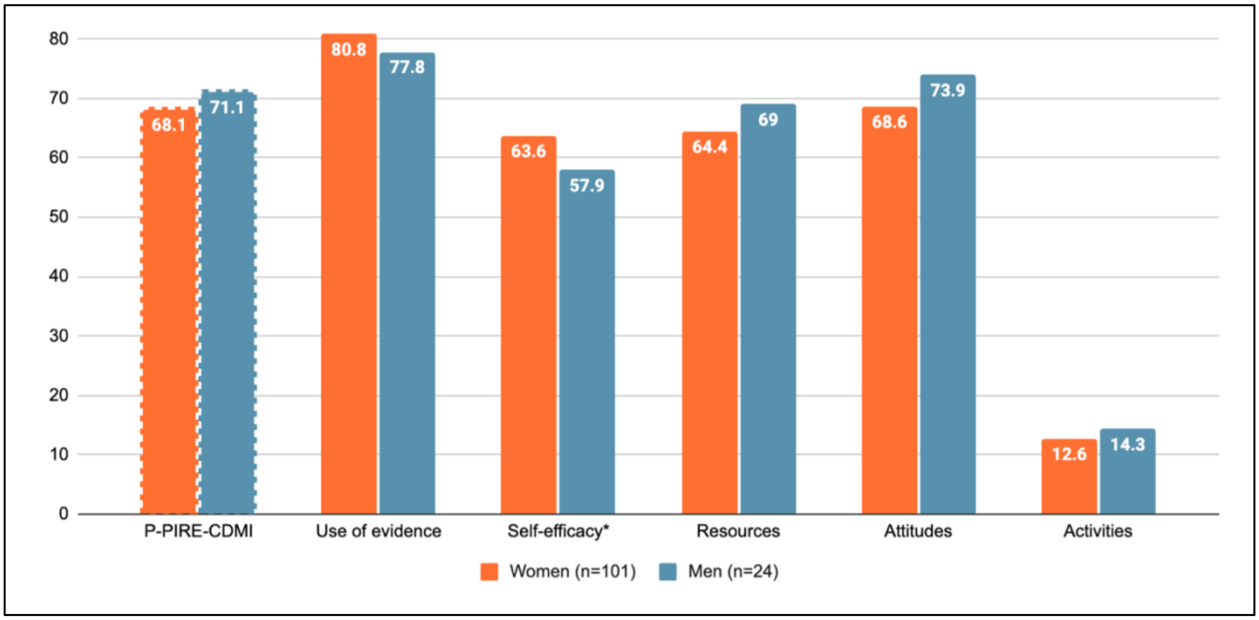


Figure 8. Distribution of scores from the prototype and five EBP measures by profession at baseline

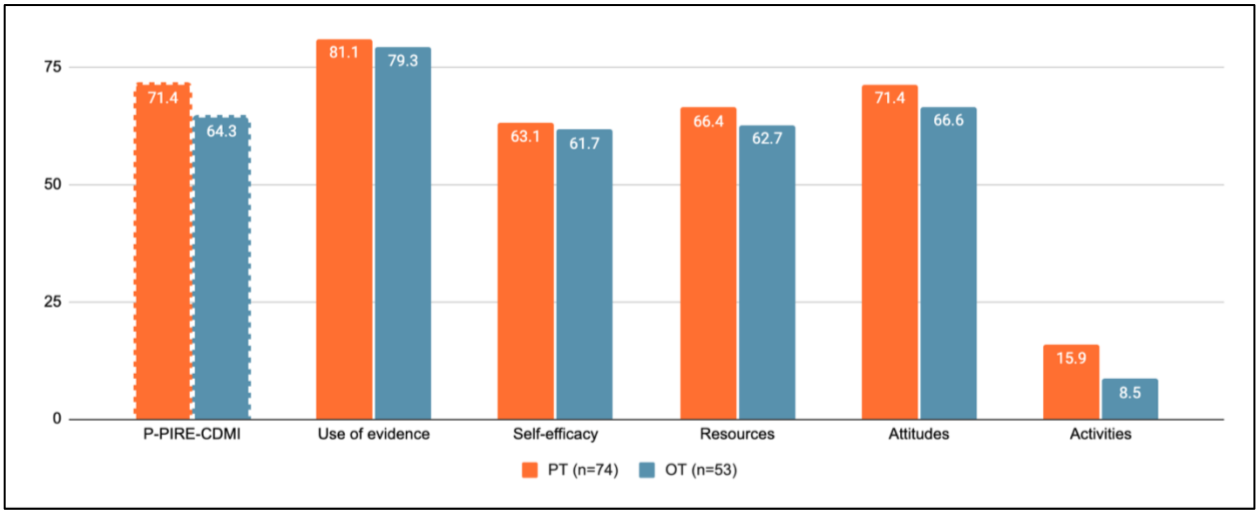
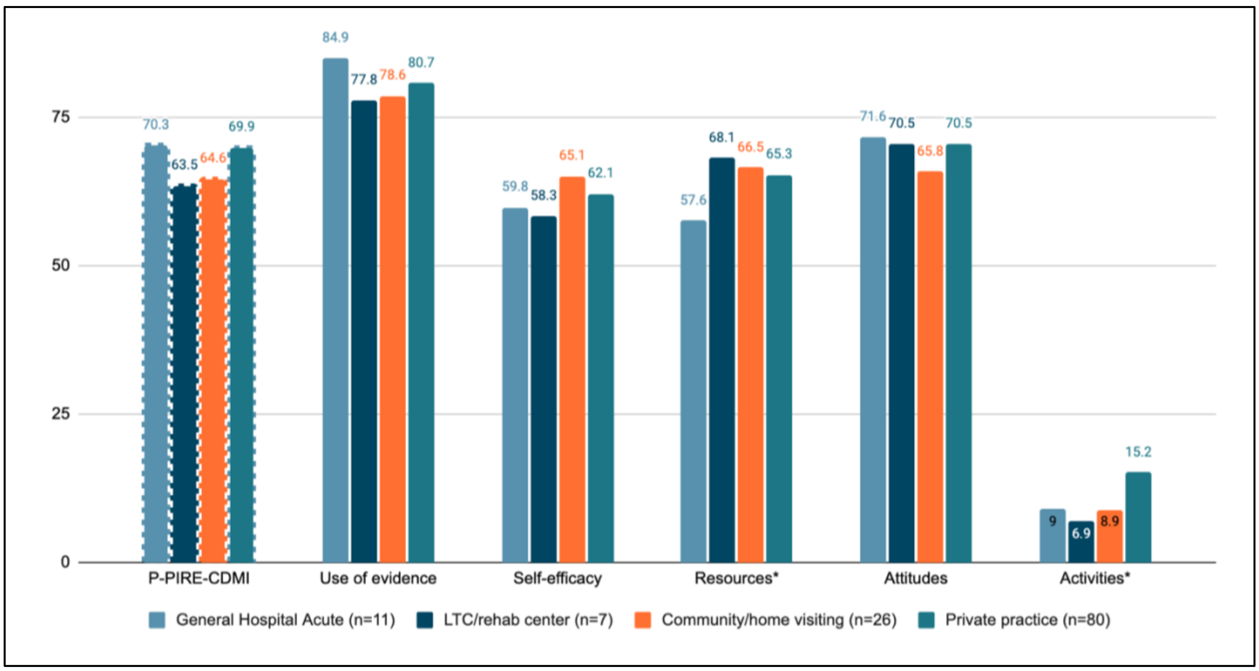


Figure 9. Distribution of scores from the prototype and five EBP measures by setting at baseline



Discussion

This study used existing survey data of newly graduated Canadian OTs and PTs [11,55] and RMT methods to identify candidate items, develop a weighted sum of dimensions and provide evidence on the interpretability of a brief prototype index of propensity to integrate research evidence into CDM. The prototype index comprises five items representative of the previously identified core EBP domains of (1) use of research evidence, (2) self-efficacy, (3) resources, (4) attitudes, and (5) activities related to EBP. Each dimension was evaluated on a three-point interval scale from the lowest response option indicating neutral propensity to the highest response option indicating higher propensity to integrate research evidence into CDM. Our results provide evidence of a multidimensional index that can provide a single score that can help rapidly identify a clinician's propensity to integrate research evidence but also highlight specific areas requiring further evaluation and support.

Polychoric correlations between selected items ranged from 0.34 to 0.57 which aligned with our a priori hypothesis. These low to moderate correlations support the posit that the items are related yet not redundant and measure separate domains of propensity to integrate research evidence into CDM[70]. The 0.57 correlation between the *attitudes* item and *resources* item was acceptable given their conceptual distance (i.e., one item relates to the organization, the other relates to the person). The expert panel omitted the *knowledge* dimension from the prototype index. This decision is justified given recent research demonstrating the absence of an association between knowledge scores and OTs' and PTs' self-reported use of research evidence[11]. In addition, challenges with the measuring the knowledge domain have been described in the literature and include low internal consistency[99], differential item functioning of knowledge items by ethnicity, gender and profession[74] and overestimation of self-reported knowledge levels[100].

The overall frequency of individuals across profiles was normally distributed for the P-PIRE-CDMI, indicating that the index may have the potential to discriminate between clinicians without any floor or ceiling effects. The P-PIRE-CDMI behaved coherently across age, gender, and settings at both time points. Also, there was no distinct source of variance on index scores across these three characteristics over time. There was an important difference in P-PIRE-CDMI scores between OTs and PTs at baseline whereby OTs scored 7 points lower out of 100 on the index than PTs. An important profession effect was also found across all EBP measures at baseline. This discrepancy between OTs and PTs at baseline may be explained by professional differences such as the scope of practice, the nature of available evidence, and the type of knowledge valued in CDM. Though we have no way of knowing with certainty why this difference exists, it is possible that some components of occupational therapy practice (e.g., environmental adaptations, occupational performance, or psychosocial approaches) may not be adequately addressed in the literature compared to components of physical therapy practice[20]. Our finding on the difference between OTs and PTs is consistent with a study conducted in the practice area of chronic pain by Arumugam et al.[9] whereby OTs scored lowest in EBP behavior compared to other healthcare professionals including PTs. The authors hypothesized that the difference may be due to the scope of practice of OTs which was more psychosocial for which there is (1) less high-level evidence and (2) evidence levels may not be the primary concern in CDM. In contrast, PTs in the area of chronic pain have roles in the biomechanical and physiological management of acute pain which is better addressed in the literature. Our results confirm the importance of stratifying results related to EBP by rehabilitation profession given that these groups exhibit different trends.

Four years later, there was no longer a difference between professions on P-PIRE-CDMI scores. PT scores dropped and OT scores remained stable. This decrease in scores may speak to the mechanism wherein clinical experience gained over time offsets the difference between professions whereby forms of knowledge (e.g., experiential knowledge) outside empirical forms become paramount[101].

Whilst some may consider this drop in scores disappointing, it is possible that the decrease in EBP over time (as measured by the P-PIRE-CDMI and the five EBP measures) can be, in part, explained by variations in the definition of evidence[102,103]. For new graduates, evidence may be closely related to research articles and textbooks[104]. With experience, the research-based knowledge once acquired through searching the literature becomes consolidated such that it may take the form of tacit or experiential knowledge and may no longer be distinguished as *evidence* or *research evidence*[104–106].

Similarly, an important age effect was found whereby P-PIRE-CDMI scores dropped on average 1.6 points for each additional year of life. Considering our minimal important difference estimate of 6.5, this trend suggests that there is a minimally important drop in propensity scores every four additional years as a rehabilitation clinician. This trend is maintained across all EBP measures with an average 0.85 drop in scores per additional year of age. Our findings are concordant with other studies that indicate negative correlations of EBP behaviors, self-efficacy and attitudes with age in physical therapists and occupational therapists [47,107–112].

The performance of the prototype index was compared to the five unidimensional EBP measures. As hypothesized, the P-PIRE-CDMI was moderately correlated with the five EBP measures. Our results suggest that the measure that was used has an important effect on overall score distribution in terms of magnitude and direction. In other words, the value on the latent construct of EBP (which these unidimensional measures are said to reflect or form) is inconsistent when using different measures. Compared to the P-PIRE-CDMI, one measure yields notably higher average scores (*use of research evidence*) and three measures yield considerably lower scores (*self-efficacy, resources,* and *activities*). A combined interpretation of results from the five unidimensional EBP measures may be thwarted due to these diverging distributions. The P-PIRE-CDMI provides a single score that combines these five key dimensions into one coherent indicator and behaves in a way that is consistent with the five EBP measures across time.

Furthermore, our findings reveal that when stratifying the sample by age, gender, setting and university, EBP measures provide very different perspectives. When considering these sociodemographic variables, the scores on the broader latent variable (EBP) produced by each measure are incoherent. Using gender as an example, the *use of research evidence* and *self-efficacy* measures suggest that men score lower whereas the three other EBP measures suggest that women score lower. We also found an important heterogeneity in the ranking of setting and university by different EBP measures. In other words, if stakeholders were to use the five EBP measures in practice to gain an overall interpretation of EBP, it would be difficult to reach an actionable conclusion to support EBP across settings and universities. For instance, the *self-efficacy* measure may direct these stakeholders to allocate resources to the lowest ranking group (long-term care/rehabilitation center), but the *resources* measure would orient these stakeholders towards focusing their support on another lowest ranking group (general hospital- acute care). Similarly, but with a larger degree of heterogeneity, it would be unfeasible to interpret scores from multiple EBP measures by university. In contexts where researchers are seeking to determine EBP engagement levels as a function of university attended or type of clinical setting, we suggest considering the use of the PIRE-CDMI (future versions) as a first brief overall measurement which may then be complemented, when necessary, with in-depth measures of specific domains of interests.

Although we do not claim that the results of individual EBP measures are inaccurate, the combined interpretation of results from these measures may be challenging and ineffective due to the opposing distributions, variation in scores based on age, gender, setting and university and difference in rank ordering of settings and universities. The P-PIRE-CDMI presents the advantage of combining five dimensions into one succinct and comprehensive measure.

Despite our best efforts, there are limitations to this study. First, we invite readers to cautiously interpret these findings as more research is required to refine these P-PIRE-CDMI scores. Although the prototype index represents a reasonable proof-of-concept measure, more research is needed before this index can be used in practice. Next steps may include qualitatively revising the items, developing a scoring algorithm which considers relative dimension weighting, and contributing validity evidence to support the use of PIRE-CDMI scores in various samples and settings (e.g., relationships to other variables, acceptability). Second, the estimates of P-PIRE-CDMI scores are based on an additive formula of the logit placements which may be mathematically valid but is not theoretically valid; to do so would require eliciting dimension weights based on the perspectives of end-users. Therefore, the use of the P-PIRE-CDMI results must be regarded as preliminary. Third, the beta estimates using the GEE methods should be interpreted with the consideration that we only used two time points. We did not aim to estimate the extent to which real change occurred; our goal was to provide evidence on the interpretability of the P-PIRE-CDMI. A group-based trajectory modeling of EBP constructs using the same dataset is published elsewhere[101].

Finally, we acknowledge that the EBP scores presented in this paper may be overestimates for three reasons. First, clinicians may provide responses in a more socially acceptable way because EBP is a professional expectation and they may not be inclined to disagree with its value (i.e., social desirability bias). Second, there is a likelihood that the convenience sample used in this paper is more positively inclined towards EBP compared to non-respondents. Thirdly, the sample used in this study is composed of new graduates who have been found to demonstrate higher knowledge[44], attitudes[44,46,109,113], skills[44,113,114] and behaviors[115] compared to established and practicing clinicians.

Given that the PIRE-CDMI will undergo continuous rigorous development and field testing in future studies, this index has numerous potential uses. In a research context, the PIRE-CDMI could be used as a clinical outcome measure to determine if other variables (e.g., clinical setting or type of education) influence propensity or to evaluate the effectiveness of interventions aimed at supporting EBP of OTs and/or PTs in clinical settings. The index could be used to enhance sampling procedures such that clinicians may be allocated to different intervention groups or included/excluded to gain a heterogeneous sample based on PIRE-CDMI scores. The PIRE-CDMI could be adopted as a self-reflection tool whereby clinicians can rapidly assess their own propensity score. Such a reflective tool could be integrated into periodic self-monitoring or could be used as a basis for communicating with managers about quality improvement. Although the next steps described above are required before the PIRE-CDMI can be employed for those purposes, this research makes important methodological and practical contributions to advancing the measurement of EBP by providing a streamlined and pragmatic approach to the measurement of a complex process that can have important consequences on health outcomes of rehabilitation patients.

Conclusion

In this study, we identified one best performing item for each of the five core EBP domains using previously conducted Rasch analysis. These five items constitute the prototype *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI) which combines five dimensions onto a single scale with a mathematically valid scoring system. Testing of the P-PIRE-CDMI demonstrates that it behaves consistently across age group, gender and setting. Compared to other unidimensional EBP measures, the P-PIRE-CDMI provides comparable information in a succinct way. This study highlights the benefits of a brief, multidimensional index to assess an OT or PT's propensity to integrate research evidence into CDM.

References

1. Leland NE, Crum K, Phipps S, Roberts P, Gage B. Advancing the Value and Quality of Occupational Therapy in Health Service Delivery. *Am J Occup Ther*. 2015;69(1):6901090010p1-6901090010p7. doi:10.5014/ajot.2015.691001

2. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1

3. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Evidence-Based Medicine E-Book: How to Practice and Teach EBM*. 5th edition. Elsevier Health Sciences; 2018.

4. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014

5. CAOT, ACOTRO, ACOTUP, PAC. Joint position statement on evidence-based occupational therapy. *Can J Occup Ther*. 2009;66(5):267-269. https://www.caot.ca/document/3697/J%20-%20Joint%20Position%20Statement%20on%20Evidence%20based%20OT.pdf

6. CCPUP. Entry-to-Practice Physiotherapy Curriculum: Content Guidelines for Canadian University Programs. Published online May 2009. https://peac-aepc.ca/pdfs/Resources/Competency%20Profiles/CCPUP%20Curriculum%20Guidelines%202019.pdf

7. CPA. Position statement on evidence-informed practice in physiotherapy. Published online 2012. https://physiotherapy.ca/sites/default/files/positionstatements/evidence-informed-practice-position-statement-english.pdf

8. Emparanza JI, Cabello JB, Burls AJE. Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *J Eval Clin Pract*. 2015;21(6):1059-1065. doi:10.1111/jep.12460

9. Arumugam V, MacDermid JC, Walton D, Grewal R. Attitudes, knowledge and behaviors related to evidence-based practice in health professionals involved in pain management. *JBI Evid Implement*. 2018;16(2):107-118. doi:10.1097/XEB.0000000000000131

10. da Silva T, da Cunha Menezes Costa L, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. 2015;20(3):388-401. doi:10.1016/j.math.2014.10.009

11. Thomas A, Zoubi FA, Mayo NE, et al. Individual and organizational factors associated with evidence-based practice among physical and occupational therapy recent graduates: A cross-sectional national study. *J Eval Clin Pract*. 2020;n/a(n/a). doi:https://doi.org/10.1111/jep.13518

12. Bérubé MÈ, Poitras S, Bastien M, Laliberté LA, Lacharité A, Gross DP. Strategies to translate knowledge related to common musculoskeletal conditions into physiotherapy practice: a systematic review. *Physiotherapy*. 2018;104(1):1-8. doi:10.1016/j.physio.2017.05.002

13. Hallé MC, Mylopoulos M, Rochette A, et al. Attributes of evidence-based occupational therapists in stroke rehabilitation. *Can J Occup Ther*. 2018;85(5):351-364. doi:10.1177/0008417418802600

14. Juckett LA, Wengerd LR, Faieta J, Griffin CE. Evidence-Based Practice Implementation in Stroke Rehabilitation: A Scoping Review of Barriers and Facilitators. *Am J Occup Ther Off Publ Am Occup Ther Assoc*. 2020;74(1):7401205050p1-7401205050p14. doi:10.5014/ajot.2020.035485

15. Myers CT, Lotz J. Practitioner Training for Use of Evidence-Based Practice in Occupational Therapy. *Occup Ther Health Care*. 2017;31(3):214-237. doi:10.1080/07380577.2017.1333183

16. Scurlock-Evans L, Upton P, Upton D. Evidence-Based Practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014;100(3):208-219. doi:10.1016/j.physio.2014.03.001

17. Salbach NM, Jaglal SB, Korner-Bitensky N, Rappolt S, Davis D. Practitioner and Organizational Barriers to Evidence-based Practice of Physical Therapists for People With Stroke. *Phys Ther*. 2007;87(10):1284-1303. doi:10.2522/ptj.20070040

18. Salbach NM, Jaglal SB, Williams JI. Reliability and validity of the Evidence-Based practice confidence (EPIC) scale. *J Contin Educ Health Prof*. 2013;33(1):33-40.

19. Munce SEP, Graham ID, Salbach NM, et al. Perspectives of health care professionals on the facilitators and barriers to the implementation of a stroke rehabilitation guidelines cluster randomized controlled trial. *BMC Health Serv Res*. 2017;17(1):440. doi:10.1186/s12913-017-2389-7

20. Robertson L, Graham F, Anderson J. What Actually Informs Practice: Occupational Therapists' Views of Evidence. *Br J Occup Ther*. 2014;76:317-324. doi:10.4276/030802213X13729279114979

21. Thomas Y, Dickson D, Broadbridge J, et al. Benefits and challenges of supervising occupational therapy fieldwork students: Supervisors' perspectives. *Aust Occup Ther J*. 2007;54:S2-S12.

22. Craik J, Rappolt S. Enhancing research utilization capacity through multifaceted professional development. *Am J Occup Ther*. 2006;60(2):155-164.

23. McQueen J, Nivison C, Husband V, Miller C. An investigation into the use of a journal club for evidence-based practice. *Int J Ther Rehabil*. 2006;13(7):311-317.

24. Stern P. Using journal clubs to promote skills for evidence-based practice. *Occup Ther Health Care*. 2008;22(4):36-53.

25. Lizarondo L, Grimmer-Somers K, Kumar S. A systematic review of the individual determinants of research evidence use in allied health. *J Multidiscip Healthc*. 2011;4:261-272. doi:10.2147/JMDH.S23144

26. McCluskey A, Vratsistas-Curto A, Schurr K. Barriers and enablers to implementing multiple stroke guideline recommendations: a qualitative study. *BMC Health Serv Res*. 2013;13(1):323. doi:10.1186/1472-6963-13-323

27. Lyons C, Brown T, Tseng MH, Casey J, McDonald R. Evidence-based practice and research utilisation: Perceived research knowledge, attitudes, practices and barriers among Australian paediatric occupational therapists. *Aust Occup Ther J*. 2011;58(3):178-186.

28. Salbach NM, Guilcher SJ, Jaglal SB, Davis DA. Determinants of research use in clinical decision making among physical therapists providing services post-stroke: a cross-sectional study. *Implement Sci*. 2010;5(1):77. doi:10.1186/1748-5908-5-77

29. Bayley MT, Hurdowar A, Richards CL, et al. Barriers to implementation of stroke rehabilitation evidence: findings from a multi-site pilot project. *Disabil Rehabil*. 2012;34(19):1633-1638. doi:10.3109/09638288.2012.656790

30. Fristedt S, Areskoug Josefsson K, Kammerlind AS. Factors Influencing the Use of Evidence-Based Practice among Physiotherapists and Occupational Therapists in Their Clinical Work. *Internet J Allied Health Sci Pract*. Published online 2016. doi:10.46743/1540-580X/2016.1553

31. Morris JH, Bernhardsson S, Bird ML, et al. Implementation in rehabilitation: a roadmap for practitioners and researchers. *Disabil Rehabil*. 2019;0(0):1-10. doi:10.1080/09638288.2019.1587013

32. Bennett S, Allen S, Caldwell E, et al. Organisational support for evidence-based practice: occupational therapists perceptions. *Aust Occup Ther J*. 2016;63(1):9-18. doi:10.1111/1440-1630.12260

33. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to Evidence-Based Practice Implementation in Physiotherapy: A Systematic Review and Meta-Analysis. *Int J Qual Health Care*. 2021;(mzab093). doi:10.1093/intqhc/mzab093

34. Roberge-Dao J, Maggio LA, Zaccagnini M, et al. Quality, methods, and recommendations of systematic reviews on measures of evidence-based practice: an umbrella review. *JBI Evid Synth*. 2022;20(4):1004-1073.

35. Mayo N. *ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*. 1st edition. International Society for Quality of Life Research; 2015.

36. Buchanan H, Siegfried N, Jelsma J. Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review: Evidence-based Practice Survey Instruments. *Occup Ther Int*. 2016;23(2):59-90. doi:10.1002/oti.1398

37. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gómez JE. Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the literature: Evidence-based physiotherapy instruments. *J Eval Clin Pract*. 2014;20(6):767-778. doi:10.1111/jep.12180

38. Stevenson K, Lewis M, Hay E. Do physiotherapists' attitudes towards evidence-based practice change as a result of an evidence-based educational programme? *J Eval Clin Pract*. 2004;10(2):207-217. doi:10.1111/j.1365-2753.2003.00479.x

39. Fernández-Domínguez JC, de Pedro-Gómez JE, Morales-Asencio JM, Bennasar-Veny M, Sastre-Fullana P, Sesé-Abad A. Health Sciences-Evidence Based Practice questionnaire (HS-EBP) for measuring transprofessional evidence-based practice: Creation, development and psychometric validation. Gagnier JJ, ed. *PLOS ONE*. 2017;12(5):e0177172. doi:10.1371/journal.pone.0177172

40. Graham F, Robertson L, Anderson J. N ew Z ealand occupational therapists' views on evidence-based practice: A replicated survey of attitudes, confidence and behaviours. *Aust Occup Ther J*. 2013;60(2):120-128.

41. Hadley J, Hassan I, Khan KS. Knowledge and beliefs concerning evidence-based practice amongst complementary and alternative medicine health care practitioners and allied health care professionals: a questionnaire survey. *BMC Complement Altern Med*. 2008;8(1):1-7.

42. Heiwe S, Kajermo KN, Tyni-Lenné R, et al. Evidence-based practice: attitudes, knowledge and behaviour among allied health care professionals. *Int J Qual Health Care*. 2011;23(2):198-209. doi:10.1093/intqhc/mzq083

43. Iles R, Davidson M. Evidence based practice: a survey of physiotherapists' current practice. *Physiother Res Int*. 2006;11(2):93-103.

44. Jette DU, Bacon K, Batty C, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. *Phys Ther*. 2003;83(9):786-805.

45. Lyons C, Casey J, Brown T, Tseng M, McDonald R. Research knowledge, attitudes, practices and barriers among paediatric occupational therapists in the United Kingdom. *Br J Occup Ther*. 2010;73(5):200-209.

46. MacEwan Dysart A, Tomlin GS. Factors Related to Evidence-Based Practice Among U.S. Occupational Therapy Clinicians. *Am J Occup Ther*. 2002;56(3):275-284. doi:10.5014/ajot.56.3.275

47. McCluskey A. Occupational therapists report a low level of knowledge, skill and involvement in evidence-based practice. *Aust Occup Ther J*. 2003;50(1):3-12. doi:10.1046/j.1440-1630.2003.00303.x

48. Nilsagård Y, Lohse G. Evidence-based physiotherapy: A survey of knowledge, behaviour, attitudes and prerequisites. *Adv Physiother*. 2010;12:179-186. doi:10.3109/14038196.2010.503812

49. Philibert DB, Snyder P, Judd D, Windsor MM. Practitioners' reading patterns, attitudes, and use of research reported in occupational therapy journals. *Am J Occup Ther*. 2003;57(4):450-458.

50. Salls J, Dolhi C, Silverman L, Hansen M. The use of evidence-based practice by occupational therapists. *Occup Ther Health Care*. 2009;23(2):134-145.

51. Stronge M, Cahill M. Self-reported Knowledge, Attitudes and Behaviour towards Evidence-based Practice of Occupational Therapy Students in Ireland: Evidence-based Practice, Irish OT students. *Occup Ther Int*. 2012;19(1):7-16. doi:10.1002/oti.328

52. McEvoy MP, Williams MT, Olds TS. Development and psychometric testing of a trans-professional evidence-based practice profile questionnaire. *Med Teach*. 2010;32(9):e373-e380.

53. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci*. 2014;9(1):118. doi:10.1186/s13012-014-0118-8

54. Glegg SMN, Holsti L. Measures of knowledge and skills for evidence-based practice: A systematic review. *Can J Occup Ther Ott*. 2010;77(4):219-232.

55. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

56. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? *J Rehabil Med*. 2012;44(2):97-98. doi:10.2340/16501977-0938

57. Jamieson S. Likert scales: How to (ab) use them? *Med Educ*. 2004;38(12):1217-1218.

58. McEvoy MP, Williams MT, Olds TS, Lewis LK, Petkov J. Evidence-based practice profiles of physiotherapists transitioning into the workforce: a study of two cohorts. *BMC Med Educ*. 2011;11(1):100.

59. Bishop PA, Herron RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int J Exerc Sci*. 2015;8(3):297-302.

60. Gupta M. Reconsidering rationality and ethics in the evidence-based medicine debate: a reply to commentators. *J Eval Clin Pract*. 2004;10(2):143-146. doi:10.1111/j.1365-2753.2004.00500.x

61. Norman GR. Examining the assumptions of evidence-based medicine: Examining the assumptions of EBM. *J Eval Clin Pract*. 1999;5(2):139-147. doi:10.1046/j.1365-2753.1999.00197.x

62. Tonelli. Integrating evidence into clinical practice: an alternative to evidence-based approaches: Integrating evidence into clinical practice. *J Eval Clin Pract*. 2006;12(3):248-256. doi:10.1111/j.1365-2753.2004.00551.x

63. Tonelli M. The philosophical limits of evidence-based medicine. Published online 1998.

64. Solomon M. Just a paradigm: evidence-based medicine in epistemological context. *Eur J Philos Sci*. 2011;1(3):451-466. doi:10.1007/s13194-011-0034-6

65. Estabrooks CA. Will evidence-based nursing practice make practice perfect? *Can J Nurs Res Arch*. Published online 1998:15-36.

66. Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. *Psychol Rev*. 2003;110(2):203-219. doi:10.1037/0033-295X.110.2.203

67. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79(387):516-524.

68. Bradburn N, Cartwright N, Fuller J. Chapter 5: A Theory of Measurement. In: *Measurement in Medicine : Philosophical Essays on Assessment and Evaluation*. ; 2017. Accessed June 23, 2022. https://web-p-ebscohost-com.proxy3.library.mcgill.ca/ehost/ebookviewer/ebook/bmxlYmtfXzE2MDkyNzJfX0FO0?sid=983c7550-88d1-4ee8-89c0-5320cfb680da@redis&vid=0&format=EB&lpid=lp_73&rid=0

69. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*. 2005;58(6):550-559. doi:10.1016/j.jclinepi.2004.10.016

70. Diamantopoulos A, Siguaw JA. Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *Br J Manag*. 2006;17(4):263-282. doi:10.1111/j.1467-8551.2006.00500.x

71. Akram W, Hussein MSE, Ahmad S, Mamat MN, Ismail NE. Validation of the knowledge, attitude and perceived practice of asthma instrument among community pharmacists using Rasch analysis. *Saudi Pharm J*. 2015;23(5):499-503. doi:10.1016/j.jsps.2015.01.011

72. Belvedere SL, de Morton NA. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *J Clin Epidemiol*. 2010;63(12):1287-1297.

73. Conrad KJ, Smith Jr EV. International conference on objective measurement: applications of Rasch analysis in health care. *Med Care*. Published online 2004:I1-I6.

74. Harris-Haywood S, Goode T, Gao Y, et al. Psychometric Evaluation of a Cultural Competency Assessment Instrument for Health Professionals: *Med Care*. 2014;52(2):e7-e15. doi:10.1097/MLR.0b013e31824df149

75. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value Health*. 2015;18(1):25-34. doi:10.1016/j.jval.2014.10.005

76. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The First Stage of Developing Preference-Based Measures: Constructing a Health-State Classification Using Rasch Analysis. *Qual Life Res*. 2009;18(2):253-265.

77. Mavranezouli I, Brazier JE, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a

measure of common mental health problems (CORE-OM). *Qual Life Res*. 2011;20(3):321-333. doi:10.1007/s11136-010-9768-4

78. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846-855. doi:10.1016/j.jval.2015.07.004

79. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. *Qual Life Res*. 2010;19(6):907-917. doi:10.1007/s11136-010-9646-0

80. Andrich D. Rasch models for ordered response categories. *Encycl Stat Behav Sci*. Published online 2005.

81. Andrich D. Rasch models. In: *International Encyclopedia of Education*. Pergamon; 2010:111-122.

82. Upton D, Upton P. Development of an evidence-based practice questionnaire for nurses. *J Adv Nurs*. 2006;53(4):454-458. doi:10.1111/j.1365-2648.2006.03739.x

83. Aarons GA, Glisson C, Hoagwood K, Kelleher K, Landsverk J, Cafri G. Psychometric properties and US National norms of the Evidence-Based Practice Attitude Scale (EBPAS). *J Psychol Assess*. 2010;22(2):356.

84. Salbach NM, Jaglal SB. Creation and validation of the evidence-based practice confidence scale for health care professionals. J Eval Clin Pract. 2011;17(4):794-800.

85. Estabrooks CA, Squires JE, Cummings GG, Birdsell JM, Norton PG. Development and assessment of the Alberta Context Tool. *BMC Health Serv Res*. 2009;9(1):234. doi:10.1186/1472-6963-9-234

86. Engelhard G. Invariant Measurement: Summary and Discussion. In: *Invariant Measurement*. Routledge; 2012.

87. Traub RE. Classical Test Theory in Historical Perspective. *Educ Meas Issues Pract*. 1997;16(4):8-14. doi:10.1111/j.1745-3992.1997.tb00603.x

88. Poissant L, Mayo NE, Wood-Dauphinee S, Clarke AE. The development and preliminary validation of a Preference-Based Stroke Index (PBSI). *Health Qual Life Outcomes*. Published online 2003:15.

89. Kuspinar A. *Development of a Preference-Based Measure for Multiple Sclerosis: The Preference-Based Multiple Sclerosis Index (PBMSI)*. McGill University; 2014. Accessed May 3, 2019. https://onlinelibrary.wiley.com/doi/abs/10.1111/scs.12482

90. Moga AM, Twells LK, Mayo NE. The prototype of a preference-based index of weight-related quality of life: demonstrating the possibilities. *Qual Life Res*. Published online May 24, 2022. doi:10.1007/s11136-022-03156-5

91. Brazier J, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32). doi:10.3310/hta16320

92. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011. doi:10.1017/CBO9780511996214

93. Diamantopoulos A, Winklhofer HM. Index Construction with Formative Indicators: An Alternative to Scale Development. *J Mark Res*. 2001;38(2):269-277.

94. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. Published online 2003:582-592.

95. Andrade C. Mean difference, standardized mean difference (SMD), and their use in meta-analysis: as simple as it gets. *J Clin Psychiatry*. 2020;81(5):11349.

96. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155.

97. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM. *J Eval Clin Pract*. 2017;23(5):964-970. doi:10.1111/jep.12752

98. Greene BA. Measuring Cognitive Engagement With Self-Report Scales: Reflections From Over 20 Years of Research. *Educ Psychol*. 2015;50(1):14-30. doi:10.1080/00461520.2014.989230

99. Shi Q, Chesworth BM, Law M, Haynes RB, MacDermid JC. A modified evidence-based practice- knowledge, attitudes, behaviour and decisions/outcomes questionnaire is valid across multiple professions involved in pain management. *BMC Med Educ*. 2014;14(1):263. doi:10.1186/s12909-014-0263-4

100. Blanch-Hartigan D. Medical students' self-assessment of performance: results from three meta-analyses. *Patient Educ Couns*. 2011;84(1):3-9.

101. Iqbal MZ, Rochette A, Mayo NE, et al. Exploring how evidence-based practice of occupational and physical therapists evolves: A longitudinal mixed methods national study. *Submitted to Plos One on July 08, 2022 (PONE-D-22-19314).*

102. Ashcroft RE. Current epistemological problems in evidence based medicine. *J Med Ethics*. 2004;30(2):131-135. doi:10.1136/jme.2003.007039

103. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic Inquiries in Evidence-Based Medicine. *Cancer Control*. 2009;16(2):158-168. doi:10.1177/107327480901600208

104. Jeffery H, Robertson L, Reay K. Sources of evidence for professional decision-making in novice occupational therapy practitioners: clinicians' perspectives. *Br J Occup Ther*. 2020;84:030802262094139. doi:10.1177/0308022620941390

105. Paterson M, Higgs J, Wilcox S. Developing expertise in judgement artistry in occupational therapy practice. *Br J Occup Ther*. 2006;69(3):115-123.

106. Paterson M, Wilcox S, Higgs J. Exploring dimensions of artistry in reflective practice. *Reflective Pract*. 2006;7(4):455-468.

107. Bennett S, Tooth L, McKenna K, et al. Perceptions of evidence-based practice: A survey of Australian occupational therapists. *Aust Occup Ther J*. 2003;50(1):13-22. doi:10.1046/j.1440-1630.2003.00341.x

108. Bridges PH, Bierema LL, Valentine T. The propensity to adopt evidence-based practice among physical therapists. *BMC Health Serv Res*. 2007;7(1):103. doi:10.1186/1472-6963-7-103

109. Cameron KAV, Ballantyne S, Msot AK, Msot M margolis-Gal, Msot TD, Ludwig F. Utilization of evidence-based practice by registered occupational therapists. *Occup Ther Int*. 2005;12(3):123-136. doi:10.1002/oti.1

110. Hitch D, Nicola-Richmond K, Richards K, Stefaniak R. Student perspectives on factors that influence the implementation of evidence-based practice in occupational therapy. *JBI Evid Implement*. 2021;19(4):409-418. doi:10.1097/XEB.0000000000000285

111. Thomas A, Law M. Research Utilization and Evidence-Based Practice in Occupational Therapy: A Scoping Study. *Am J Occup Ther*. 2013;67(4):e55-e65. doi:10.5014/ajot.2013.006395

112. Wang Y, King K, Moran B, et al. Occupational and Physical Therapists' Perception of Evidence-Based Practice. *J Allied Health Wash*. 2019;48(2):119-126,126A-126C.

113. de Pedro Gomez J, Morales-Asencio JM, Abad AS, Veny MB, Roman MJ, Ronda FM. Validation of the Spanish version of the evidence based practice questionnaire in nurses. *Rev Esp Salud Publica*. 2009;83(4):577-586.

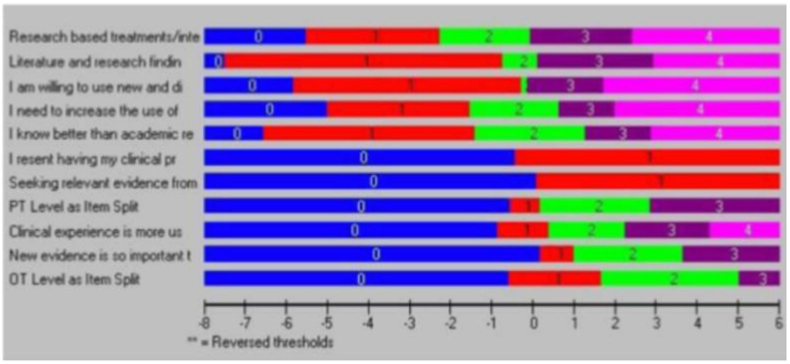
114. Upton P, Scurlock-Evans L, Stephens D, Upton D. The adoption and implementation of evidence-based practice (EBP) among allied health professions. *Int J Ther Rehabil*. 2012;19(9):497-503.

115. Prior P, Wilkinson J, Neville S. Practice nurse use of evidence in clinical practice: a descriptive survey. *Nurs Prax N Z*. 2010;26(2).

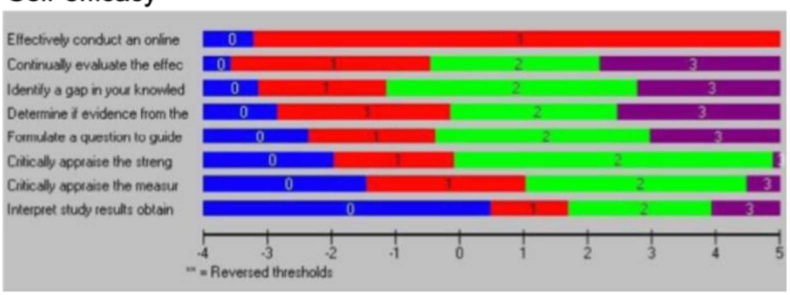
Appendices

Appendix I. Rasch threshold maps for the reflective constructs developed by Al Zoubi et al.[55]. Authorization to use figures received from the core research team.

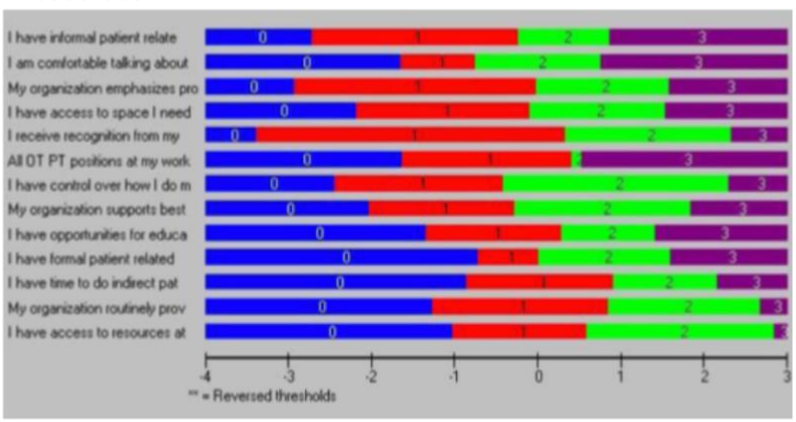
Attitudes



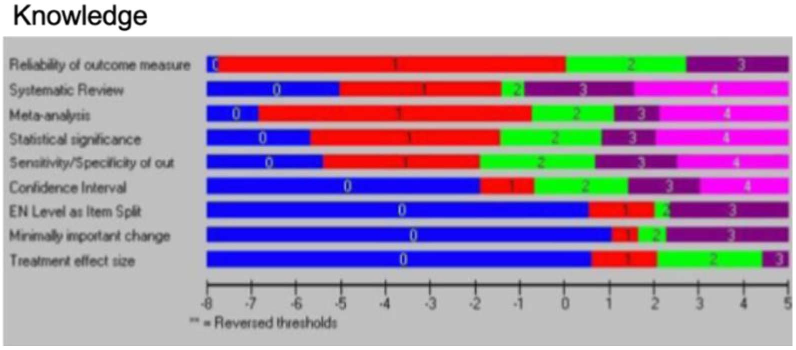
Self-efficacy



Resources



Knowledge



Appendix II: Prototype PIRE-CDMI inter-item polychoric correlation matrix

	Use of research evidence	Self-efficacy	Resources	Attitudes
Self-efficacy	0.17			
Resources	0.34	0.39		
Attitudes	0.26	0.35	0.57	
Activities	0.12	0.34	0.30	0.47

Appendix III. Distribution of scores of EBP measures across characteristics of the sample at T0 (N=127) and at T3 (n=37)

	Measures of EBP at T0											
Characteristics	**P-PIRE-CDMI**		**Use of research evidence**		**Self-efficacy**		**Resources**		**Attitudes**		**Activities**	
	5 items		9 items		8 items		13 items		10 items		7 items	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AGE												
25 and below (n=44)	69.7	13.8	80.6	17.1	62.9	17.7	65.9	13.6	69.3	12.3	12	11.8
26 to 27 (n=49)	68.1	12.8	80.2	16.4	62.9	16.9	64.1	13.7	67.9	12.2	12.9	13.1
28 and above (n=34)	67.3	11.8	80.1	17.7	61.3	18.2	64.7	16.9	71.6	10.2	13.6	15.9
GENDER [a]												
Men (n=24)	71.1	13.8	77.8	16.7	57.9	13.8	69	14.8	73.9	14	14.3	14.6
Women (n=101)	68.1	12.5	80.8	17	63.6	18.2	64.4	13.8	68.6	10.7	12.6	13.3
PROFESSION												
Physical therapist (n=74)	71.4	11.9	81.1	16.8	63.1	16.9	66.4	13.4	71.4	11	15 .9	15.4
Occupational therapist (n=53)	64.3	13	79.3	17	61.7	18.3	62.7	15.8	66.6	12.2	8.5	8.5
SETTING [b]												
General Hospital - acute care (n=11)	70.3	11.3	84.9	13.4	59.8	21.5	57.6	13.5	71.6	8.6	9	6.7
Long term care or rehabilitation center (n=7)	63.5	18.9	77.8	19.3	58.3	27.2	68.1	18.4	70.5	16.2	6.9	5.9
Community agency, primary health care, home visiting	64.6	11.1	78.6	18.6	65.1	19.7	66.5	14.1	65.8	12.6	8.9	13.4

agency, consulting firm (n=26)												
Private practice (n=80)	69.9	13	80.7	16.6	62.1	15.1	65.3	14.6	70.5	11.2	15.2	14.3
UNIVERSITY												
B (n=10)	72.1	12	74.4	17.4	61.8	16.6	60.8	11.4	70.6	9	14.3	13
C (n=18)	70.2	13.6	87	18.4	75.2	19.5	60.5	11.2	68.1	12.2	17.6	16.4
D (n=13)	74.1	10.9	86.3	13.7	61.5	17.3	69.4	11.5	72.6	11.1	16	15.1
E (n=9)	72	13.3	85.2	13.6	49.5	11.5	65.2	15.8	70.5	12.5	13.7	15
F (n=20)	68	10.9	76.7	18.4	66.6	15.3	61.5	14.7	69.2	14.1	15.1	16.3
G (n=7)	73.3	7.2	84.1	18	55.2	16.7	67	13.7	77.2	9.7	18.9	19.2
H (n=14)	68.3	13.5	81	16.6	56.5	12.4	71.3	11.8	71	7	9	6
I (n=5)	68.4	12.3	86.7	14.5	55.5	13.4	71.8	11.8	70.6	13	13.4	12
J (n=22)	60.2	14.7	71.2	15.6	63.4	20.1	66.3	20.6	64.1	12.2	4.6	3.2
Group of 6 universities [c] (n=9)	66.7	11.5	82.7	13.7	60.1	13.2	59.5	11.8	69.1	12.7	11.2	10.5
TIME												
T0 (n=127)	68.4	12.9	80.3	16.9	62.5	17.4	64.9	14.5	69.4	11.7	12.8	13.4
T0 with follow-up at T3 (n=37)	67.8	12.8	77.8	17.4	67.5	20.1	63.3	16.3	70.2	9	12	12
T3 (n=37)	64.2	11.9	71.8	15.4	61.8	15.1	63.8	13.7	63.2	9.7	7.9	7
Difference T3-T0 (n=37)	-3.7	14.3	-6	19.9	-5.7	14.5	0.4	17.6	-7	12.3	-4.1	10.4
SRM [d]	-0.26		-0.3		-0.39		0.02		-0.6		-0.39	
Paired t-test (95% CI, df: 36)	-1.57		-1.84		-2.38		0.14		-3.47		-2.4	
RANGE [e] (min, max, range)												
T0 (n=127)	42	95.1	55.6	100	13.6	99.1	20.5	94.9	37.5	96.9	0	75

T0 with follow-up at T3 (n=37)	42.6	95.1	55.6	100	22.7	99.1	20.5	94.9	53.1	96.9	0.7	50
T3 (n=37)	41.4	92.9	55.6	100	31.8	95.5	35.9	94.9	43.8	87.5	0.7	25

[a] No reporting for the *did not specify* group because n=2 (less than 5)

[b] No reporting for the *Missing or N/A* group because n=3 (less than 5)

[c] Six universities were grouped together (A, K, L, M, N, O) because sample sizes were less than 5 per university

[d] SRM is a metric of effect size

[e] The theoretical range for each measure is 0-100.

EBP, evidence-based practice; P-PIRE-CDMI**,** prototype of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index;* SD, standard deviation; SRM, standardized response mean calculated as (T3- T0)/SD of the group's score differences; CI, confidence interval; df, degrees of freedom

Appendix IV. Regression estimates of the effect of characteristic on P-PIRE-CDMI scores at T0 and T3

P-PIRE-CDMI score = characteristic	T0				T3			
	Unstandardized Coefficients		95% CI for ß		Unstandardized Coefficients		95% CI for ß	
	Estimate (ß)	SE	Lower Bound	Upper Bound	Estimate (ß)	SE	Lower Bound	Upper Bound
AGE- ***CATEGORICAL***								
*25 and below (n_{T0}=44; n_{T3}=16)	69.7	1.9	65.9	73.6	63.6	3.1	57.4	69.8
26,27 (n_{T0}=49; n_{T3}=11)	-1.6	2.7	-6.9	3.7	2.0	4.8	-7.7	11.8
28 above (n_{T0}=34; n_{T3}=10)	-2.4	2.9	-8.3	3.4	-0.2	4.9	-10.2	9.8
3 outliers excluded								
*25 and below (n_{T0}=44; n_{T3}=16)	69.7	1.9	65.9	73.5	n=16	63.6	<.001	58.0
26,27 (n_{T0}=49; n_{T3}=11)	-1.6	2.6	-6.8	3.6	n=11	2.0	0.639	-6.7
28 above (n_{T0}=31; n_{T3}=7)	-0.7	3.0	-6.6	5.2	n=7	-0.3	0.945	-10.5
AGE - ***CONTINOUS***								
Constant	91.2	11.4	68.6	113.9		75.5	<.001	47.0
Age at Time 0 (n_{T0}=127; n_{T3}=37)	-0.8	0.4	-1.7	0.0	n=37	-0.4	0.421	-1.5
3 outliers excluded								
Constant	114.4	30.7	51.9	176.8		79.9	0.008	22.2
Age at Time 0 (n_{T0}=124; n_{T3}=34)	-1.7	1.2	-4.1	0.7	n=34	-0.6	0.582	-2.8
GENDER								
*Men (n_{T0}=24; n_{T3}=4)	71.1	2.6	66.0	76.3	62.9	5.6	51.5	74.3
Women (n_{T0}=101; n_{T3}=32)	-3.0	2.9	-8.7	2.7	0.5	5.9	-11.6	12.6
PROFESSION								
*Physical therapist (n_{T0}=74; n_{T3}=15)	71.4	1.4	68.5	74.2	64.6	3.1	58.3	71.0

Occupational therapist (n_{T0}=53; n_{T3}=22)	-7.1	2.2	-11.5	-2.6	-0.8	4.0	-9.0	7.4
SETTING								
*Private practice (n_{T0}=80; n_{T3}=18)	69.9	1.4	67.1	72.8	64.5	2.9	58.6	70.4
General Hospital Acute (n_{T0}=11; n_{T3}=6)	0.4	4.1	-7.8	8.5	-2.9	5.8	-14.7	8.9
Long term care/rehabilitation center (n_{T0}=7; n_{T3}=1)	-6.4	5.1	-16.4	3.6				
Community agency, primary health care, home visiting agency, consulting firm (n_{T0}=26; n_{T3}=10)	-5.3	2.9	-11.0	0.4	1.7	4.8	-8.2	11.6

*referent category

P-PIRE-CDMI**,** prototype of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index; ß,* beta; SE, standard error; CI, confidence interval

Appendix V: GEE estimates of the effects of characteristic and time on P-PIRE-CDMI scores

1) Age categorical with and without outliers (group 1 = 25 and below; group 2 = 26 and 27; group 3 = 28 and above)

P-PIRE-CDMI =age (categorical) + time + interaction (Age group*Time)

Parameter Estimates (with outliers, n=37)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	71.3	3.2	64.9	77.7	482.3	1	0
[time=2]	-7.7	2.9	-13.3	-2.1	7.1	1	0.008
[time=1]	0a	.	.	.	.	.	.
[age_group=3]	-11.0	4.4	-19.6	-2.4	6.2	1	0.012
[age_group=2]	-1.7	4.8	-11.2	7.7	0.1	1	0.721
[age_group=1]	0a	.	.	.	.	.	.
[time=2] * [age_group=3]	10.8	6.1	-1.1	22.7	3.2	1	0.076
[time=2] * [age_group=2]	3.7	4.7	-5.4	12.9	0.6	1	0.421
[time=2] * [age_group=1]	0a	.	.	.	.	.	.
[time=1] * [age_group=3]	0a	.	.	.	.	.	.
[time=1] * [age_group=2]	0a	.	.	.	.	.	.
[time=1] * [age_group=1]	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100							

Parameter Estimates (without outliers, n=34)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	71.3	3.2	64.9	77.7	482.3	1	0
[time=2]	-7.7	2.9	-13.3	-2.1	7.1	1	0.008
[time=1]	0a	.	.	.	.	.	.
[age_group=3]	-6.3	3.9	-13.9	1.3	2.7	1	0.102
[age_group=2]	-1.7	4.8	-11.2	7.7	0.1	1	0.721
[age_group=1]	0a	.	.	.	.	.	.
[time=2] * [age_group=3]	6.0	4.0	-1.8	13.8	2.2	1	0.134
[time=2] * [age_group=2]	3.7	4.7	-5.4	12.9	0.6	1	0.421
[time=2] * [age_group=1]	0a	.	.	.	.	.	.
[time=1] * [age_group=3]	0a	.	.	.	.	.	.
[time=1] * [age_group=2]	0a	.	.	.	.	.	.
[time=1] * [age_group=1]	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100							

2) Age continuous with and without outliers

P-PIRE-CDMI =age (continuous) + time + interaction (Age *Time)

Parameter Estimates (with outliers, n=37)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	111.0	9.9	91.6	130.3	126.5	1	0

[time=2]	-35.5	24.7	-83.8	12.8	2.1	1	0.15
[time=1]	0a	.	.	.	.	.	.
Age at T0	-1.6	0.3	-2.2	-0.9	23.2	1	<.001
[time=2] * Age at Time 0	1.2	0.9	-0.7	3.0	1.5	1	0.213
[time=1] * Age at Time 0	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100							

Parameter Estimates (<u>without</u> outliers, n=34)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	114.4	23.6	68.1	160.6	23.5	1	0
[time=2]	-34.5	20.8	-75.2	6.2	2.8	1	0.097
[time=1]	0a	.	.	.	.	.	.
Age at T0	-1.7	0.9	-3.4	0.0	3.9	1	0.048
[time=2] * Age at Time 0	1.1	0.8	-0.4	2.7	2.1	1	0.152
[time=1] * Age at Time 0	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100							

<u>3) Gender</u>

P-PIRE-CDMI =gender (binary) + time + interaction (Gender*Time)

Parameter Estimates (n=35)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	72.6	6.4	60.2	85.1	130.4	1	0
[time=2]	-9.7	5.2	-20.0	0.5	3.5	1	0.063
[time=1]	0a	.	.	.	.	.	.
[women]	-4.6	6.7	-17.8	8.5	0.5	1	0.489
[men]	0a	.	.	.	.	.	.
[time=2] * [women]	5.1	5.6	-5.8	16.0	0.8	1	0.358
[time=2] * [men]	0a	.	.	.	.	.	.
[time=1] * [women]	0a	.	.	.	.	.	.
[time=1] * [men]	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100 N.B. “Did not specify group” not included because n=2 at T0							

<u>4) Profession</u>

P-PIRE-CDMI =profession (binary) + time + interaction (Profession*Time)

Parameter Estimates (n=37)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	72.7	2.5	67.7	77.6	826.6	1	0
[time=2]	-8.1	2.5	-13.0	-3.1	10.3	1	0.001
[time=1]	0a	.	.	.	.	.	.
[OT]	-8.2	3.8	-15.6	-0.7	4.7	1	0.031
[PT]	0a	.	.	.	.	.	.
[time=2] * [OT]	7.4	4.2	-0.8	15.6	3.1	1	0.078
[time=2] * [PT]	0a	.	.	.	.	.	.

[time=1] * [OT]	0a	.	.	.	.	.	.
[time=1] * [PT]	0a	.	.	.	.	.	.
Dependent Variable: P-PIRE-CDMI T0 100							

<u>5) Setting</u>

P-PIRE-CDMI =setting + time + interaction (setting*Time)

Parameter Estimates (n=37)							
Parameter	Estimate (ß)	Standard error	95% Wald Confidence Interval		Wald Chi-Square	df	Sig.
			Lower	Upper			
(Intercept)	66.6	3.3	60.2	73.0	417.3	1.0	0
[time=2]	-5.0	4.9	-14.5	4.6	1.0	1.0	0.309
[time=1]	0a	.	.	.	.	.	.
[Missing or N/A setting]	1.7	5.7	-9.4	12.8	0.1	1.0	0.763
[Private practice]	2.0	4.8	-7.4	11.3	0.2	1.0	0.677
[Community agency, primary health care, home visiting agency, consulting firm]	-1.2	4.7	-10.3	7.9	0.1	1.0	0.797
[Long term care or rehabilitation center]	19.1	3.3	12.8	25.5	34.5	1.0	0
[Acute care]	0a	.	.	.	.	.	.
[time=2] * [Missing or N/A setting]	-8.2	15.1	-37.9	21.5	0.3	1.0	0.589
[time=2] * [Private practice]	0.9	6.1	-11.1	12.8	0.0	1.0	0.886
[time=2] * [Community agency, primary health care, home visiting agency, consulting firm]	5.8	5.7	-5.3	16.8	1.0	1.0	0.309
[time=2] * [Long term care or rehabilitation center]	-9.3	4.9	-18.9	0.2	3.7	1.0	0.055
[time=2] * [Acute care]	0a	.	.	.	.	.	.
[time=1] * [Missing or N/A]	0a						
[time=1] * [Private practice]	0a						
[time=1] * [Community agency, primary health care, home visiting agency, consulting firm]	0a						
[time=1] * [Long term care or rehabilitation center]	0a						
[time=1] * [Acute care]	0a						
Dependent Variable: P-PIRE-CDMI T0 100 N.B. Missing or N/A group n=2 at T3; Long-term care or rehab center n=1 at T3							

Appendix VI. GEE estimates of measure; measure and characteristic; and measure and time on scores

1. SCORE = MEASURE

A GEE model was used to estimate the effects of measure on score, accounting for the correlation between multiple measures on the same individual.

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	68.4	1.1	66.2	70.7	3630.7	1	.000
[Activities measure]	-55.7	1.2	-57.9	-53.4	2312.7	1	.000
[Resources measure]	-3.6	1.4	-6.2	-0.9	6.9	1	.009
[Use of research evidence measure]	11.9	1.4	9.1	14.7	70.2	1	.000
[Self-efficacy measure]	-5.9	1.5	-9.0	-2.9	14.7	1	.000
[Attitudes measure]	0.9	1.0	-1.1	3.0	0.8	1	.362
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
(Scale)	213.6						
Dependent Variable: Score on each measure Model: (Intercept), Measure							
a. Set to zero because this parameter is redundant.							

2. SCORE = MEASURE + AGE (C) + MEASURE*AGE (C)

A GEE model was used to estimate the effects of measure, age (centered) and the interaction between measure age, accounting for the correlation between multiple measures on the same individual.

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	68.4	1.1	66.2	70.6	3747.2	1	.000
[Activities measure]	-55.7	1.1	-57.9	-53.4	2385.8	1	.000
[Resources measure]	-3.6	1.4	-6.2	-0.9	6.9	1	.009
[Use of research evidence measure]	11.9	1.4	9.1	14.7	70.3	1	.000
[Self-efficacy measure]	-5.9	1.5	-8.9	-2.9	15.0	1	.000
[Attitudes measure]	1.0	1.0	-1.0	2.9	0.9	1	.339
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
Age centered	-0.8	0.4	-1.7	0.0	4.2	1	.041
[Activities measure activities] * Age centered	0.9	0.3	0.2	1.5	6.4	1	.012
[Resources measure] * Age centered	0.4	0.6	-0.7	1.5	0.5	1	.478
[Use of research evidence measure] * Age centered	0.1	0.5	-0.8	1.0	0.0	1	.876
[Self-efficacy measure] * Age centered	1.0	0.6	-0.3	2.2	2.4	1	.124
[Attitudes measure attitudes] * Age centered	1.1	0.3	0.5	1.7	13.7	1	.000

[P-PIRE-CDMI] * Age centered	0[a]	.	.	.	.	.	.
(Scale)	213.4						
Dependent Variable: Score on each measure Model: (Intercept), Measure, Age centered, Measure * Age centered							
a. Set to zero because this parameter is redundant.							

3. SCORE = MEASURE + GENDER (3 groups) + MEASURE*GENDER

A GEE model was used to estimate the effects of measure, gender (3 groups) and the interaction between measure gender, accounting for the correlation between multiple measures on the same individual.

**N.B. analysis with gender without the 'did not specify' group follows*

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	71.1	2.8	65.7	76.5	663.7	1	.000
[Activities measure]	-56.9	3.0	-62.8	-50.9	353.5	1	.000
[Resources measure]	-2.1	3.2	-8.3	4.1	0.4	1	.508
[Use of research evidence measure]	6.7	3.0	0.7	12.6	4.9	1	.028
[Self-efficacy measure]	-13.2	3.5	-20.0	-6.3	14.2	1	.000
[Attitudes measure]	2.8	2.3	-1.6	7.3	1.6	1	.210
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
[Prefer not to specify]	-18.4	7.7	-33.5	-3.3	5.7	1	.017
[Women]	-3.0	3.0	-8.9	2.9	1.0	1	.321
[Men]	0[a]	.	.	.	.	.	.
[Activities measure] * [Prefer not to specify]	9.5	7.6	-5.3	24.3	1.6	1	.208

[Activities measure] * [Women]	1.3	3.3	-5.1	7.7	0.2	1	.686
[Activities measure] * [Men]	0[a]	.	.	.	.	.	.
[Resources measure] * [Prefer not to specify]	-10.9	7.2	-24.9	3.2	2.3	1	.129
[Resources measure] * [Women]	-1.6	3.5	-8.5	5.3	0.2	1	.643
[Resources measure] * [Men]	0[a]	.	.	.	.	.	.
[Use of research evidence measure] * [Prefer not to specify]	29.5	3.1	23.4	35.6	90.7	1	.000
[Use of research evidence measure] * [Women]	6.0	3.4	-0.7	12.7	3.1	1	.080
[Use of research evidence measure] * [Men]	0[a]	.	.	.	.	.	.
[Self-efficacy measure] * [Prefer not to specify]	21.8	15.6	-8.8	52.4	2.0	1	.162
[Self-efficacy measure] * [Women]	8.7	3.9	1.1	16.2	5.0	1	.026
[Self-efficacy measure] * [Men]	0[a]	.	.	.	.	.	.
[Attitudes measure] * [Prefer not to specify]	-0.9	15.1	-30.4	28.7	0.0	1	.954
[Attitudes measure] * [Women]	-2.4	2.5	-7.3	2.6	0.9	1	.353
[Attitudes measure] * [Men]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Prefer not to specify]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Women]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Men]	0[a]	.	.	.	.	.	.
(Scale)	211.0						
Dependent Variable: Score on each measure Model: (Intercept), Measure, Gender, Measure * Gender							
a. Set to zero because this parameter is redundant.							

4. SCORE = MEASURE + GENDER (binary) + MEASURE*GENDER (binary)

A GEE model was used to estimate the effects of measure, gender (2 groups) and the interaction between measure gender, accounting for the correlation between multiple measures on the same individual.

Model Information

Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	125
	Within-Subject Effect	Measure	6
Number of Subjects			125
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates

Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	71.1	2.8	65.7	76.5	663.7	1	.000
[Activities measure]	-56.9	3.0	-62.8	-50.9	353.5	1	.000
[Resources measure]	-2.1	3.2	-8.3	4.1	0.4	1	.508
[Use of research evidence measure]	6.7	3.0	0.7	12.6	4.9	1	.028
[Self-efficacy measure]	-13.2	3.5	-20.0	-6.3	14.2	1	.000
[Attitudes measure]	2.8	2.3	-1.6	7.3	1.6	1	.210
[P-PIRE-CDMI]	0a	.	.	.	.	.	.
[Women]	-3.0	3.0	-8.9	2.9	1.0	1	.321
[Men]	0a	.	.	.	.	.	.
[Activities measure] * [Women]	1.3	3.3	-5.1	7.7	0.2	1	.686
[Activities measure] * [Men]	0a	.	.	.	.	.	.
[Resources measure] * [Women]	-1.6	3.5	-8.5	5.3	0.2	1	.643
[Resources measure] * [Men]	0a	.	.	.	.	.	.
[Use of research evidence measure] * [Women]	6.0	3.4	-0.7	12.7	3.1	1	.080
[Use of research evidence measure] * [Men]	0a	.	.	.	.	.	.
[Self-efficacy measure] * [Women]	8.7	3.9	1.1	16.2	5.0	1	.026
[Self-efficacy measure] * [Men]	0a	.	.	.	.	.	.
[Attitudes measure] * [Women]	-2.4	2.5	-7.3	2.6	0.9	1	.353
[Attitudes measure] * [Men]	0a	.	.	.	.	.	.

[P-PIRE-CDMI] * [Women]	0a	.	.	.	.	.	.
[P-PIRE-CDMI] * [Men]	0a	.	.	.	.	.	.
(Scale)	210.4						
Dependent Variable: Score on each measure Model: (Intercept), Measure, genderbinary, Measure * genderbinary							
a. Set to zero because this parameter is redundant.							

5. SCORE = MEASURE + PROFESSION + MEASURE* PROFESSION

A GEE model was used to estimate the effects of measure, profession (2 groups) and the interaction between measure profession, accounting for the correlation between multiple measures on the same individual.

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	71.4	1.4	68.7	74.1	2677.3	1	.000
[Activities measure]	-55.5	1.7	-58.8	-52.2	1084.8	1	.000
[Resources measure]	-5.0	1.8	-8.5	-1.4	7.4	1	.007
[Use of research evidence measure]	9.7	1.8	6.3	13.1	30.6	1	.000
[Self-efficacy measure]	-8.3	1.8	-11.8	-4.8	21.6	1	.000
[Attitudes measure]	0.0	1.3	-2.5	2.5	0.0	1	.989
[P-PIRE-CDMI]	0a	.	.	.	.	.	.
[OT]	-7.1	2.2	-11.5	-2.7	9.9	1	.002
[PT]	0a	.	.	.	.	.	.
[Activities measure] * [OT]	-0.3	2.2	-4.7	4.0	0.0	1	.876

[Activities measure] * [PT]	0a	.	.	.	.	.	.
[Resources measure] * [OT]	3.3	2.7	-2.0	8.7	1.5	1	.221
[Resources measure] * [PT]	0a	.	.	.	.	.	.
[Use of research evidence measure] * [OT]	5.2	2.9	-0.4	10.9	3.3	1	.070
[Use of research evidence measure] * [PT]	0a	.	.	.	.	.	.
[Self-efficacy measure] * [OT]	5.7	3.2	-0.6	12.0	3.1	1	.077
[Self-efficacy measure] * [PT]	0a	.	.	.	.	.	.
[Attitudes measure] * [OT]	2.3	2.1	-1.9	6.5	1.2	1	.280
[Attitudes measure] * [PT]	0a	.	.	.	.	.	.
[P-PIRE-CDMI] * [OT]	0a	.	.	.	.	.	.
[P-PIRE-CDMI] * [PT]	0a	.	.	.	.	.	.
(Scale)	209.3						
Dependent Variable: Score on each measure Model: (Intercept), Measure, Profession, Measure * Profession							
Profession = 0 : PT Profession = 1 : OT							

6. SCORE = MEASURE + SETTING + MEASURE* SETTING

A GEE model was used to estimate the effects of measure, setting (5 groups) and the interaction between measure setting, accounting for the correlation between multiple measures on the same individual.

General Hospital - acute care (n=11)
Long term care or rehabilitation center (n=7)
Community agency, primary health care, home visiting agency, consulting firm (n=26)
Private practice (n=80)
Missing or N/A (n-3)

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	70.3	3.2	63.9	76.6	470.1	1	.000
[Activities measure]	-61.3	2.9	-66.9	-55.7	462.5	1	.000
[Resources measure]	-12.7	3.5	-19.5	-5.9	13.6	1	.000
[Use of research evidence measure]	14.6	3.2	8.2	20.9	20.1	1	.000
[Self-efficacy measure]	-10.4	5.1	-20.4	-0.5	4.3	1	.039
[Attitudes measure]	1.3	3.6	-5.8	8.4	0.1	1	.719
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
[Missing or N/A]	-3.8	4.7	-13.1	5.4	0.7	1	.420
[Private practice]	-0.4	3.6	-7.3	6.6	0.0	1	.921
[Community, etc.]	-5.7	3.9	-13.3	1.9	2.1	1	.144
[Long-term care or rehab center]	-6.8	7.4	-21.2	7.7	0.8	1	.358
[Hospital - acute care]	0[a]	.	.	.	.	.	.
[Activities measure] * [Missing or N/A]	5.3	5.9	-6.2	16.8	0.8	1	.365
[Activities measure activities] * [Private practice]	6.5	3.3	0.1	13.0	4.0	1	.046
[Activities measure] * [Community, etc.]	5.6	3.4	-1.0	12.2	2.8	1	.095
[Activities measure] * [Long-term care or rehab center]	4.8	5.6	-6.2	15.7	0.7	1	.395
[Activities measure] * [Hospital - acute care]	0a	.	.	.	.	.	.
[Resources measure] * [Missing or N/A]	4.4	7.3	-9.9	18.6	0.4	1	.549
[Resources measure resources] * [Private practice]	8.1	3.8	0.6	15.6	4.5	1	.034
[Resources measure resources] * [Community, etc.]	14.6	4.5	5.8	23.3	10.6	1	.001
[Resources measure resources] * [Long-term care or rehab center]	17.3	9.2	-0.6	35.3	3.6	1	.058
[Resources measure] * [Hospital - acute care]	0a	.	.	.	.	.	.
[Use of research evidence measure] * [Missing or N/A]	-7.0	12.5	-31.5	17.5	0.3	1	.577
[Use of research evidence measure] * [Private practice]	-3.8	3.8	-11.2	3.6	1.0	1	.311
[Use of research evidence measure] * [Community, etc.]	-0.6	4.3	-9.0	7.9	0.0	1	.898
[Use of research evidence measure] * [Long-term care or rehab center]	-0.3	6.1	-12.3	11.7	0.0	1	.962
[Use of research evidence measure] * [Hospital - acute care]	0a	.	.	.	.	.	.
[Self-efficacy measure] * [Missing or N/A]	15.2	11.5	-7.3	37.7	1.8	1	.186

[Self-efficacy measure] * [Private practice]	2.6	5.4	-8.0	13.1	0.2	1	.635
[Self-efficacy measure] * [Community, etc.]	11.0	6.1	-1.1	23.0	3.2	1	.074
[Self-efficacy measure] * [Long-term care or rehab center]	5.3	8.9	-12.2	22.8	0.3	1	.556
[Self-efficacy measure] * [Hospital - acute care]	0[a]	.	.	.	.	.	.
[Attitudes measure] * [Missing or N/A]	-7.4	7.9	-22.9	8.1	0.9	1	.351
[Attitudes measure] * [Private practice]	-0.7	3.8	-8.3	6.8	0.0	1	.850
[Attitudes measure] * [Community, etc.]	-0.2	4.4	-8.8	8.4	0.0	1	.966
[Attitudes measure] * [Long-term care or rehab center]	5.7	4.7	-3.5	15.0	1.5	1	.226
[Attitudes measure] * [Hospital - acute care]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Missing or N/A]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Private practice]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Community, etc.]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Long-term care or rehab center]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Hospital - acute care]	0[a]	.	.	.	.	.	.
(Scale)	214.1						
Dependent Variable: Score on each measure Model: (Intercept), Measure, New_setting_regrouped, Measure * New_setting_regrouped							
a. Set to zero because this parameter is redundant.							

7. SCORE = MEASURE + UNI + MEASURE* UNI

A GEE model was used to estimate the effects of measure, university (10 groups, anonymized) and the interaction between measure university, accounting for the correlation between multiple measures on the same individual.

Model Information		
Dependent Variable		Score on each measure
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	127
	Within-Subject Effect	Measure	6
Number of Subjects			127
Number of Measurements per Subject	Minimum		6
	Maximum		6
Correlation Matrix Dimension			6

Parameter Estimates				
Parameter	ß	Std. Error	95% Wald Confidence Interval	Hypothesis Test

			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	66.7	3.6	59.7	73.8	341.3	1	.000
[Activities measure]	-55.5	3.4	-62.3	-48.8	259.9	1	.000
[Resources measure]	-7.2	3.2	-13.5	-0.9	5.0	1	.026
[Use of research evidence measure]	16.0	5.8	4.7	27.3	7.7	1	.006
[Self-efficacy measure]	-6.6	2.2	-11.0	-2.3	8.9	1	.003
[Attitudes measure]	2.4	3.2	-3.9	8.6	0.5	1	.459
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
[University J]	-6.5	4.7	-15.8	2.7	1.9	1	.166
[University I]	1.7	6.1	-10.3	13.6	0.1	1	.785
[University H]	1.6	5.0	-8.2	11.5	0.1	1	.747
[University G]	6.6	4.4	-2.0	15.3	2.2	1	.134
[University F]	1.2	4.3	-7.2	9.7	0.1	1	.777
[University E]	5.3	5.5	-5.5	16.1	0.9	1	.339
[University D]	7.4	4.6	-1.7	16.4	2.5	1	.112
[University C]	3.5	4.8	-5.8	12.9	0.5	1	.462
[University B]	5.3	5.1	-4.7	15.3	1.1	1	.296
[Group of 6 universities]	0[a]	.	.	.	.	.	.
[Activities measure] * [University J]	0.0	4.4	-8.6	8.6	0.0	1	1.000
[Activities measure] * [University I]	0.6	4.4	-8.1	9.2	0.0	1	.898
[Activities measure] * [University H]	-3.6	4.6	-12.5	5.4	0.6	1	.433
[Activities measure] * [University G]	1.1	8.3	-15.1	17.3	0.0	1	.897
[Activities measure] * [University F]	2.7	5.0	-7.2	12.5	0.3	1	.598
[Activities measure] * [University E]	-2.8	4.9	-12.4	6.8	0.3	1	.565
[Activities measure] * [University D]	-2.6	5.2	-12.8	7.7	0.2	1	.626
[Activities measure] * [University C]	2.9	4.3	-5.5	11.3	0.5	1	.501
[Activities measure] * [University B]	-2.2	4.1	-10.3	5.9	0.3	1	.589
[Activities measure] * [Group of 6 universities]	0[a]	.	.	.	.	.	.
[Resources measure] * [University J]	13.3	5.0	3.4	23.2	7.0	1	.008
[Resources measure] * [University I]	10.6	6.0	-1.2	22.3	3.1	1	.077
[Resources measure] * [University H]	10.1	5.6	-0.8	21.0	3.3	1	.070
[Resources measure] * [University G]	0.9	5.5	-9.9	11.6	0.0	1	.874
[Resources measure] * [University F]	0.8	4.8	-8.7	10.2	0.0	1	.873
[Resources measure] * [University E]	0.4	4.5	-8.3	9.2	0.0	1	.924
[Resources measure] * [University D]	2.5	4.9	-7.1	12.1	0.3	1	.606

[Resources measure] * [University C]	-2.5	4.3	-10.9	5.8	0.3	1	.555
[Resources measure] * [University B]	-4.1	4.4	-12.8	4.6	0.9	1	.354
[Resources measure] * [Group of 6 universities]	0[a]	.	.	.	.	.	.
[Use of research evidence measure] * [University J]	-5.0	6.7	-18.2	8.3	0.5	1	.463
[Use of research evidence measure] * [University I]	2.3	8.2	-13.8	18.4	0.1	1	.782
[Use of research evidence measure] * [University H]	-3.4	7.3	-17.6	10.8	0.2	1	.641
[Use of research evidence measure] * [University G]	-5.2	7.0	-19.0	8.6	0.5	1	.459
[Use of research evidence measure] * [University F]	-7.3	6.9	-20.9	6.3	1.1	1	.295
[Use of research evidence measure] * [University E]	-2.8	7.5	-17.6	12.0	0.1	1	.710
[Use of research evidence measure] * [University D]	-3.7	6.7	-16.9	9.4	0.3	1	.575
[Use of research evidence measure] * [University C]	0.8	6.9	-12.7	14.3	0.0	1	.906
[Use of research evidence measure] * [University B]	-13.6	7.3	-27.8	0.6	3.5	1	.061
[Use of research evidence measure] * [Group of 6 universities]	0a	.	.	.	.	.	.
[Self-efficacy measure] * [University J]	9.8	4.8	0.4	19.2	4.2	1	.040
[Self-efficacy measure] * [University I]	-6.3	5.2	-16.5	3.8	1.5	1	.223
[Self-efficacy measure] * [University H]	-5.2	3.0	-11.0	0.6	3.1	1	.077
[Self-efficacy measure] * [University G]	-11.5	5.0	-21.3	-1.8	5.3	1	.021
[Self-efficacy measure] * [University F]	5.3	4.4	-3.3	13.8	1.5	1	.228
[Self-efficacy measure] * [University E]	-15.9	6.1	-27.8	-3.9	6.8	1	.009
[Self-efficacy measure] * [University D]	-5.9	4.8	-15.4	3.5	1.5	1	.219
[Self-efficacy measure] * [University C]	11.5	4.1	3.6	19.5	8.0	1	.005
[Self-efficacy measure] * [University B]	-3.6	5.8	-14.9	7.7	0.4	1	.531

[Self-efficacy measure] * [Group of 6 universities]	0[a]	.	.	.	.	.	.
[Attitudes measure] * [University J]	1.5	4.0	-6.3	9.3	0.1	1	.705
[Attitudes measure] * [University I]	-0.1	5.4	-10.7	10.4	0.0	1	.979
[Attitudes measure] * [University H]	0.3	4.5	-8.6	9.1	0.0	1	.953
[Attitudes measure] * [University G]	1.5	5.0	-8.3	11.3	0.1	1	.762
[Attitudes measure] * [University F]	-1.1	4.5	-9.9	7.6	0.1	1	.804
[Attitudes measure] * [University E]	-3.9	4.4	-12.5	4.7	0.8	1	.376
[Attitudes measure] * [University D]	-3.9	4.5	-12.8	5.1	0.7	1	.396
[Attitudes measure] * [University C]	-4.6	4.1	-12.6	3.5	1.2	1	.269
[Attitudes measure] * [University B]	-3.8	5.0	-13.6	6.0	0.6	1	.446
[Attitudes measure] * [Group of 6 universities]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University J]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University I]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University H]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University G]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University F]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University E]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University D]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University C]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [University B]	0[a]	.	.	.	.	.	.
[P-PIRE-CDMI] * [Group of 6 universities]	0[a]	.	.	.	.	.	.
(Scale)	203.4						

Dependent Variable: Score on each measure
Model: (Intercept), Measure, university_new, Measure * university_new

a. Set to zero because this parameter is redundant.

Group of 6 universities with sample sizes <5 (A, K, L, M, N, O) (n=9)
B (n=10)
C (n=18)
D (n=13)
E (n=9)
F (n=20)
G (n=7)

H (n=14)
I (n=5)
J (n=22)

8. Score=measure + time + measure*time

A GEE model was used to estimate the effects of measure, time (T0, T3) and the interaction between measure time, accounting for the correlation between multiple measures on the same individual at both timepoints.

Model Information		
Dependent Variable		Score on EBP measures
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
	2	Time
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	37
	Within-Subject Effect	Measure	6
		Time	2
Number of Subjects			37
Number of Measurements per Subject	Minimum		12
	Maximum		12
Correlation Matrix Dimension			12

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	67.8	2.1	63.8	71.9	1067.7	1	.000
[Activities measure]	-55.8	1.7	-59.0	-52.6	1140.5	1.00	0.00
[Resources measure]	-4.5	2.7	-9.8	0.8	2.8	1.00	0.10
[Use of research evidence measure]	10.0	2.4	5.2	14.7	16.7	1.00	0.00
[Self-efficacy measure]	-0.4	3.6	-7.3	6.6	0.0	1.00	0.92
[Attitudes measure]	2.4	2.3	-2.1	6.8	1.1	1.00	0.30
[P-PIRE-CDMI]	0a	.	.	.	.	.	.
[Time=2]	-3.7	2.3	-8.2	0.9	2.5	1.00	0.11
[Time=1]	0a	.	.	.	.	.	.
[Activities measure] * [Time=2]	-0.4	2.2	-4.8	4.0	0.0	1.00	0.85
[Activities measure] * [Time=1]	0a	.	.	.	.	.	.
[Resources measure] * [Time=2]	4.1	3.1	-2.0	10.1	1.8	1.00	0.19

[Resources measure] * [Time=1]	0a	.	.	.	.	.	.
[Use of research evidence measure] * [Time=2]	-2.3	3.3	-8.9	4.2	0.5	1.00	0.48
[Use of research evidence measure] * [Time=1]	0a	.	.	.	.	.	.
[Self-efficacy measure] * [Time=2]	-2.0	2.9	-7.8	3.7	0.5	1.00	0.49
[Self-efficacy measure] * [Time=1]	0a	.	.	.	.	.	.
[Attitudes measure] * [Time=2]	-3.3	3.0	-9.2	2.5	1.3	1.00	0.26
[Attitudes measure] * [Time=1]	0a	.	.	.	.	.	.
[P-PIRE-CDMI] * [Time=2]	0a	.	.	.	.	.	.
[P-PIRE-CDMI] * [Time=1]	0a	.	.	.	.	.	.
(Scale)	191.5						
Dependent Variable: Score on EBP measures Model: (Intercept), Measure, Time, Measure * Time							
a. Set to zero because this parameter is redundant.							

9. Score=measure + time

Model Information		
Dependent Variable		Score on EBP measures
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	ID
Within-Subject Effect	1	Measure
	2	Time
Working Correlation Matrix Structure		Exchangeable

Correlated Data Summary			
Number of Levels	Subject Effect	ID	37
	Within-Subject Effect	Measure	6
		Time	2
Number of Subjects			37
Number of Measurements per Subject	Minimum		12
	Maximum		12
Correlation Matrix Dimension			12

Parameter Estimates							
Parameter	ß	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	68.2	1.9	64.5	71.8	1356.1	1.0	0.0
[Activities measure]	-56.0	1.2	-58.3	-53.7	2261.7	1.0	0.0

[Resources measure]	-2.4	2.2	-6.7	1.8	1.3	1.0	0.3
[Use of research evidence measure]	8.8	1.8	5.4	12.2	25.1	1.0	0.0
[Self-efficacy measure]	-1.4	2.5	-6.3	3.6	0.3	1.0	0.6
[Attitudes measure]	0.7	1.7	-2.6	3.9	0.2	1.0	0.7
[P-PIRE-CDMI]	0[a]	.	.	.	.	.	.
[Time=2]	-4.3	1.2	-6.7	-2.0	12.9	1.0	0.0
[Time=1]	0[a]	.	.	.	.	.	.
(Scale)	190.8						
Dependent Variable: Score on EBP measures Model: (Intercept), Measure, Time							
a. Set to zero because this parameter is redundant.							

Chapter 6: The Integration of Manuscripts 3 and 4

Research objectives of manuscripts 3 and 4

Manuscript 3:

The global aim of this study was to describe the prototype development of the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific study objectives were to: (1) identify candidate items best reflecting the most salient EBP domains for OTs and PTs across Canada from existing measures; (2) estimate the extent to which the prototype PIRE-CDMI "behaves" coherently across characteristics of the sample; and (3) estimate the extent to which the prototype PIRE-CDMI provides information that is comparable to that from other EBP measures.

Manuscript 4:

The global aim of this study was to contribute evidence for the clarity and interpretability of items and response options for a new bilingual (English and French) measure, *the Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). Specifically, the primary objective of this research was to qualitatively review and revise the included items in the prototype index in English and French The secondary objective was to estimate the equivalency of response option labels in English and French.

Integration of manuscripts 3 and 4

Manuscript 3 describes the initial steps taken to develop the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). Expanding on research by Al Zoubi and colleagues[1], an initial pool of 70 items covering six EBP domains (knowledge, use of research evidence, self-efficacy, resources, attitudes, and activities related to EBP) was reduced to five items, one item covering each dimensions (minus the knowledge dimension) The item reduction methods consisted of previously conducted Rasch analysis and expert consensus[2]. After the five items were selected for inclusion in the prototype index, a scoring algorithm was developed using preliminary weights based on the logit placements on the Rasch continuum. This allowed for a single score to be generated to examine the interpretability of the prototype index score across characteristics of the sample of OTs and PTs (from a previous dataset[1]) and compared to other EBP measures. This study highlights the benefits of this brief index and provides sources of evidence related to content, internal structure and relationships to other variables supporting the continuation of the PIRE-CDMI development process.

While *Manuscript* 3 aligned with phases *a* and *b* of the four phases of measure development reported in *The Standards*[3], Manuscript *4* builds on phase *b*. Phase *a* involves establishing the purpose, intended users, construct and content inclusion of the measure while phase *b* involves developing, testing and evaluating items. In the following manuscript, *Manuscript 4*, items included in the prototype index were reviewed and rewritten iteratively for clarity and interpretability by rehabilitation clinicians and experts in the field of EBP. This was a necessary step, in part, because the five selected items were derived from various questionnaires resulting in differences between items and response options in terms of the terminology and formulation which can increase respondent burden and introduce measurement bias. Further, the English and French versions of these items may present cultural or linguistic discrepancies that can further introduce systematic differences in scores[4]. Following the focus group and cognitive interviews described in *Manuscript 4*, important modifications were made to the items and response options. The simultaneous translation approach was used in the focus group as an efficient method for addressing linguistic discrepancies. During the cognitive interviews, modifications to items suggested in one language led to the equivalent modification in the other language. Revised items were then reviewed in subsequent cognitive interviews. The equivalency of response options in both languages was verified through a cross-sectional survey of native English and French speakers.

References

1. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

2. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. *Qual Life Res*. 2010;19(6):907-917. doi:10.1007/s11136-010-9646-0

3. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing*.; 2014.

4. McGorry SY. Measurement in a cross-cultural environment: survey translation issues. *Qual Mark Res Int J*. Published online 2000.

Chapter 7: Manuscript 4

Improving Clarity and Interpretability of Items in a Bilingual Index of Propensity to Integrate Research Evidence into Clinical Decision-Making in Rehabilitation

Jacqueline Roberge-Dao[1,2], Nancy Mayo[1,3], Annie Rochette[2,4], Keiko Shikako[1,2], Aliki Thomas[1,2,5]

[1] School of Physical and Occupational Therapy, McGill University, Montréal, Canada

[2] Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Canada

[3]Center for Outcomes Research and Evaluation (CORE), The Research Institute of the McGill University Health Center (RI-MUHC), Montreal, QC, Canada

[4] L'École de réadaptation de l'Université de Montréal, Montréal, Canada

[5] The Institute of Health Sciences Education, McGill University, Montréal, Canada

Abstract

Objectives: The aim of this study was to contribute evidence for the clarity and interpretability of items and response options for a new bilingual (English and French) measure, the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI).

Methods: This study was conducted in three phases: (1) An online focus group with rehabilitation clinicians and experts in the field of EBP was conducted to review the items and response options of the PIRE-CDMI for clarity, consistency, and interval properties and arrive at a consensus on modifications that would be needed to have equivalent versions in English and French. (2) This was followed by cognitive interviews using the verbal probing method with occupational and physical therapists in English and French. Participants were asked to elaborate on their interpretation on the meaning of the item, the comprehensibility of items, and the appropriateness of response options; they were also asked to suggest alternate wording. Cognitive interviews were stopped when no further changes to items were necessary as suggested by three consecutive participants. (3) A cross-sectional online survey was conducted to validate the English and French equivalency of response options on a 0 to 100 scale for the five included items. Multiple linear regression was conducted to estimate the effect of response option and language on PIRE-CDMI scores.

Results: The focus group included seven participants and resulted in major modifications to the initial items of the prototype index. Cognitive interviews (n=24) were conducted with 14 occupational therapists and 10 physical therapists in English and French. The PIRE-CDMI underwent 12 iterations of changes overall with substantial modifications to the *use of research evidence* and *attitudes* items. No important main effects of language or interaction effects of language with response option were found on 0-100 scale ratings.

Conclusion: The qualitative revision process allowed for clarification of items and instructions, and harmonization of items in terms of their formulation and response options. This research increases the clinical relevance and reduces measurement error of the PIRE-CDMI. The results support the equivalence of the PIRE-CDMI response options in English and French.

Introduction

To enhance quality of care, occupational (OTs) and physical therapists (PTs) are expected to engage in evidence-based practice (EBP), that is, integrate best available research evidence, their clinical expertise, and patient values and preferences when making clinical decisions[1–4]. This tripartite conceptualization of EBP has traditionally been depicted by the "three circles" model[5–8]. In more recent years, there have been refinements and additions to this conceptualization of EBP reflected in the inclusion of the organizational context to highlight the external frame which also influences clinical decision-making (CDM)[9–12].

Rehabilitation clinicians acknowledge that clinical experience is essential to integrating research evidence into practice[13–15] and as such, they must make sense of the quality, pertinence and applicability of research evidence using their judgment and tacit knowledge[16,17]. Importantly, patient-centered practice is considered a basic tenet of occupational and physical therapy and accommodating patients' goals, values and preferences are vital for positive patient outcomes and satisfaction[18,19]. Patients are encouraged to be active participants in CDM related to the entire rehabilitation process including goals, expectations, assessments, and treatments.

Despite the purported benefits of such a CDM approach and the implementation of EBP content into entry level OT and PT curricula[1,4], rehabilitation clinicians continue to report difficulties with the integration of research evidence into practice[20–25]. A lack of allotted time for activities related to EBP; poor access to formal sources of evidence; low confidence in applying research to practice; lack of knowledge that evidence-based interventions exist; and inadequate equipment to implement new practices are some of the main reasons for the underutilization of research evidence in practice[15,20–22,25–27].

Robust measurement practices are needed to identify the factors related to EBP that should be improved or the strong areas that must be maintained[28–30]. Identifying which areas require improvement can inform targeted allocation of resources to support EBP and ultimately, improve health outcomes. There exists a vast selection of questionnaires measuring the core factors influencing an OT or PT's likelihood to integrate research evidence such as their self-efficacy towards applying research to practice or the available resources to support EBP[28,31–33]. However, there are shortcomings to current EBP measures including the failure to concurrently measure multiple EBP domains, the inappropriate analysis of items derived from ordinal scales and the unknown relative weight of EBP domains.

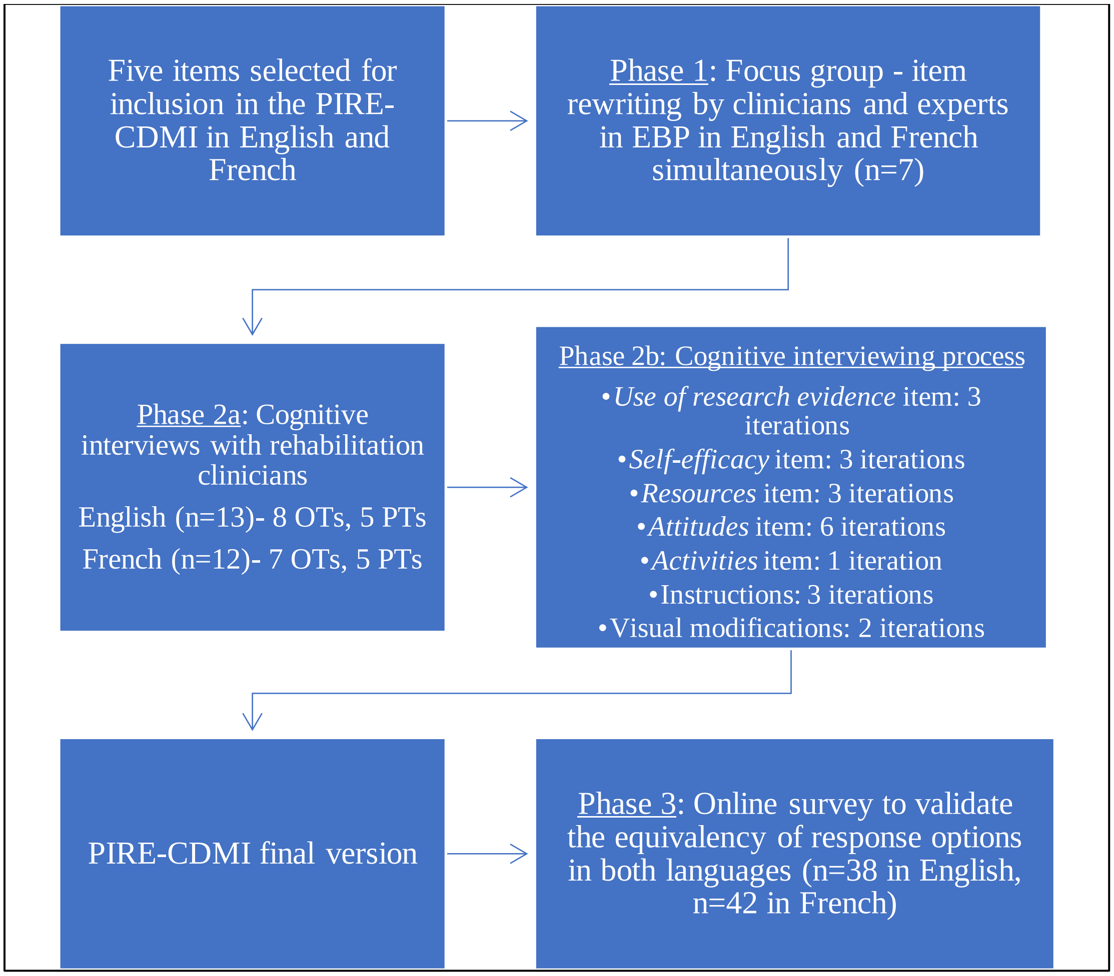
In previous work, Al Zoubi et al. identified the six most salient domains influencing a rehabilitation clinician's likelihood of integrating research evidence into CDM[30]. Then, one best performing item was chosen per EBP domain to form a brief, multidimensional index, as described in chapter 5 of this dissertation. The research reported in this manuscript, which builds on and extends previous measure development work[30], responds to a need for (1) a comprehensive measure spanning the key personal and organizational EBP domains, and (2) an efficient measurement approach to increase feasibility of use in clinical contexts and acceptability by clinicians. The intended purpose of this bilingual index is to identify needs related to the integration of research evidence in practice. Specifically, the index measures a rehabilitation clinician's propensity or likelihood to integrate research evidence into CDM. The five domains included in the index are: *use of research evidence, self-efficacy, resources, attitudes,* and *activities related to EBP*. However, as the selected items stem from five different questionnaires, there is inconsistency between items (and response options) in terms of the terminology and formulation which can increase respondent burden and introduce measurement bias. In addition, the English and French versions of these items may present cultural or linguistic discrepancies that can further introduce systematic differences in scores[34].

The global aim of this study was to contribute evidence for the clarity and interpretability of items and response options for a new bilingual (English and French) measure, the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). Specifically, the primary objective of this research was to qualitatively review and revise the included items in the prototype index in English and French The secondary objective was to estimate the equivalency of response option labels in both languages.

Methods

This study involves a three-phased qualitative review process as illustrated in Figure 1. Ethics approval was obtained from *The Faculty of Medicine and Health Sciences Institutional Review Board* at McGill University for all phases of this study before commencement and recruitment.

Figure 1. Overview of the item revision process for the Propensity to Integrate Research Evidence into Clinical Decision-Making Index (PIRE-CDMI)



Phase 1: Focus group

In measure development, focus groups with end-users and/or other stakeholders are used to (1) identify important themes to measure, and (2) discuss if selected items are relevant, clear, unambiguous, and written in terms that are understood by potential respondents[35,36].

Focus group participants

Practicing rehabilitation clinicians (OTs and PTs) and experts in EBP (defined as having experience in EBP research and having published a minimum of one EBP-related publication) were recruited purposefully from the networks of the research team to participate in a 90-minute online focus group. The pool of participants was expected to be bilingual and have equal representation of both French and English native speakers. Participants were compensated for their time with a 15$ gift card.

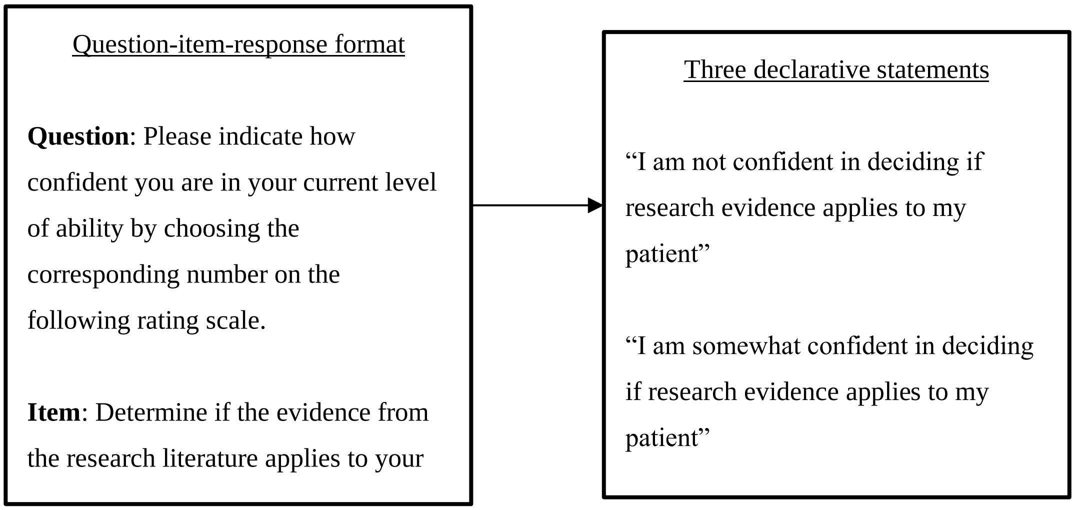
Focus group process

The aim of this focus group was to review the items and response options of the prototype PIRE-CDMI for clarity, consistency, and interval properties and arrive at a consensus on modifications that would be needed to have equivalent versions in English and French. Participants were asked to establish equivalence in both languages such that the items, instructions, and response options were conceptually (i.e., do people in both groups see the concept in the same way) and semantically (i.e., the meaning attached to words in an item) comparable[36,37]. Consenting participants were sent the items with a reminder of the study aim a week before the focus group.

The online focus group was conducted and recorded on Zoom. The structure of the session was as follows: (1) introduction and overview of the study; (2) tour de table; (3) objectives, instructions, and an example for item rewriting; (4) breakout room with two individuals per room for five minutes to allow attempts at reviewing one item; (5) attend to any questions that arose during breakout room; (6) item rewriting exercise using the share screen function.

During the item rewriting exercise, the moderator structured the discussion and a note taker recorded suggested modifications on a shared document. Participants were asked to rewrite items from question-item format into declarative statements from the perspective of a clinician (see Figure 2 for an example of a declarative statement). The moderator probed participants on the following three points: (1) How would you rewrite this item into a declarative statement? (2) Is the wording clear, and if not, how would you change it? (3) How difficult would it be for OTs and PTs to answer these items? For each item, the French translation was discussed simultaneously. Once every item was discussed, the moderator asked participants to verify that the overall index was coherent in terms of wording and length, that items read well together and that everyone agreed on the final set of items. After the focus group, the research team (consisting of bilingual EBP expert researchers in rehabilitation) reviewed the suggested final set of items and resolved any withstanding discrepancies.

Figure 2. Example of the item rewriting exercise in the focus group



Phase 2: Cognitive interviews

Cognitive interviews are used when developing measures to elicit respondents' thought process and determine if individuals (1) understand the items as intended and (2) believe the items to be relevant and appropriate[36,38,39].

Cognitive interview participants

Clinicians were eligible to participate in the cognitive interviews if they were (1) practicing OTs and PTs in Canada; (2) native French or English speaking; and (3) had been practicing for a minimum of one year. The recruitment strategy consisted of using social media (Twitter and Facebook) and McGill's School of Physical and Occupational Therapy newsletter email to advertise the project. Interested participants entered their contact information in an online form which allowed a member of the research team to contact them and provide more information on the study. Participants were compensated for their time with a 10$ gift card.

Cognitive interview process

Cognitive interviews were conducted with potential respondents of the PIRE-CDMI (i.e., OTs and PTs) to identify and rewrite any problematic items to increase the overall readability, functioning and interpretability of the measure[40–42]. Interviews were conducted by the first author, a bilingual OT and doctoral candidate with experience in conducting interviews and content knowledge in measurement. Interviews were conducted by telephone or by Zoom conferencing based on the preference of the participant and were audio recorded. Each interview was expected to last between 15 to 30 minutes.

Participants were provided with a copy of the newly reviewed PIRE-CDMI at least one day before the scheduled interview and had the option of reading the items before the interview. As presented in Table 1, the interviewer used the verbal probing method to elicit participants' comprehension of all five items by asking specific questions regarding meaning, clarity, and interpretation of items[40,43]. These questions were adapted from the authors of a study using similar methods in developing a preference-based index for multiple sclerosis[44]. Participants were encouraged to think out loud while going through the measure, allowing for insight into how a participant perceived and interpreted the items [43].

Table 1: Cognitive interview probing questions

Items
What does this statement mean to you? In your own words, what do you think this statement is saying? Were these statements easy to understand? Are there any words in this statement that are not clear or do not work well? How would you change the wording to make it clearer?
Response options
What do you think about the three options? How would you make the three options clearer?
Overall impression of the measure
Do you have any comments on the measure as a whole? Is there anything that you would change in the measure? Would you change anything with the visual presentation?

Analysis of cognitive interviews

English and French interviews were conducted in parallel so that no language was prioritized. The scheduling of interviews was based on participants' availability. After each day of interviews, which comprised between 2 to 4 individual interviews, the interviewer (JRD) reviewed comments and revised the problematic items based on the participants' suggestions. Members of the research team, which included five researchers with expertise in the development of measures and/or EBP in rehabilitation, reviewed the feedback after each day of interviews before implementing the change. The research team proposed suggestions based on best practices of item development such as having simple items that express a single idea, using common vocabulary, and avoiding colloquialisms[36,41,45]. Changes were implemented in both languages simultaneously, when applicable. The revised version of the PIRE-CDMI was then tested on the next round of participants. Given that it is recommended to continue interviewing participants until no new concerns are identified[46], interviews were conducted until no further changes were necessary as suggested by three consecutive participants[44].

Phase 3: Survey

Given Canada's linguistic diversity, methods to ensure the equivalence of questionnaire versions in different languages are warranted to decrease systematic differences between language groups. Validation of translations through quantitative response scaling can contribute to measurement equivalence by demonstrating that respondents are interpreting items in a similar fashion. Response scaling methods that have been used in previous studies[47,48] consist of asking respondents to denote the position of response options on a visual analogue scale (VAS) (i.e., a line from 0-100) and to compare the ratings between languages.

Survey respondents

This phase consisted of a cross-sectional online survey to generate additional evidence on the equivalency of PIRE-CDMI response option labels in English and French. The target population was healthcare professionals and students provided that (1) they were native English or French speakers and (2) worked or studied in a healthcare professional or graduate program in Canada. The recruitment strategy consisted of using social media and newsletter emails. We used convenience sampling and did not exclude respondents based on profession or level of training because the nature of the survey was such that respondents solely needed to have the abilities to interpret common words and rate response option labels on a numerical scale. Interested respondents were invited to follow a link with study information, a consent statement, and an invitation to start the survey in the native language of their choice, if eligible. No identifying nor sociodemographic information data were collected. Respondents did not receive any form of compensation for their time.

Survey procedure and analysis

The survey was piloted with seven graduate students, all of whom were also practicing clinicians in rehabilitation. Based on their feedback, modifications were integrated to improve survey clarity and task comprehension. The survey was open from November to December 2021.

Each PIRE-CDMI item was associated with three response option labels. Respondents were asked to indicate the position of each of the three response option labels on a 0-100 VAS between two anchors. Response option labels belonging to the same set appeared on a single page sequentially. This method has been previously reported for health-related quality of life measures such as the SF-36[47] and EuroQol-5d[48]. Specifically, participants were asked: "On the line, where would you position each of the three response option labels between [the bottom anchor] and [the top anchor]?" Appendix I presents the three response option labels for each item and associated response anchors. The full PIRE-CDMI item was also stated on the same page as the response set to provide the respondent with context.

The first page of the survey consisted of a description of the study aim, a standard set of instructions and one example which showed logical placements of the words *not very confident, somewhat confident,* and *confident* between the anchors *no confidence* and *full confidence* in relation to an example item not included in the PIRE-CDMI ("I am ___ in my ability to conduct an online literature search").

It was hypothesized that respondents would position the lowest response option label closest to the zero anchor, the middle response option label in the middle, and the highest response option label closest to the 100 anchor. Post hoc analyses were conducted to remove respondents who likely misunderstood the task reflected in having either: (1) two or more sets of disordinal response patterns; (2) the same rating for all three response options for one item or more; or (3) two or more extreme outlier ratings (0 or 100). Multiple linear regression was used to estimate the extent to which VAS ratings per item (0 to 100 scale) depended on language (English, French), response option (low, middle, high) and the interaction between language and response option. The normal probability plot of standardized residuals was visually examined.

Results

Focus group item rewriting

Four PTs and three OTs, all doctoral candidates with research experience in EBP in rehabilitation, participated in the 90-minute focus group. During the focus group, each of the five question-item-response sets was transformed into five sets of three declarative statements which were then simplified, clarified, and harmonized in English and French (version 2 of the PIRE-CDMI). Participants agreed that response options including the word "never" (for example, "I never integrate research evidence") were perceived as being undesirable because choosing these response options would make them seem incompetent. Given that clinicians would not opt for these options, the participants removed the word "never" from the *use of research evidence* and *activities* items. For the *activities* item, focus group participants communicated that it was important not to confine research evidence to scientific articles; they suggested replacing "read research" with "consult research evidence". Further, participants agreed that omitting the verb "reading" was more inclusive to individuals who may have visual impairments. In French, multiple terms were proposed for "research evidence" (e.g., évidences ou preuves scientifiques, données issues de la recherche) but the agreed upon term was *données probantes* which was said to be most employed and recognized among clinicians.

Some issues remained unresolved after the focus group period and were subsequently discussed within the research team. Modifications were made to the items before starting the cognitive interviews (version 3). First, it was unclear which term was preferred between "patient" and "client" as the terms are often used interchangeably depending on the setting and the population. The research team agreed to use "patient" consistently and added a footnote to explain the interchangeable nature (N.B. the final wording of items did not include the words "patient" nor "client"). Second, participants could not decide which term was best between "organization" and "clinical setting". The research team modified the item to focus on the broad availability of resources and refrained from using either term. Finally, participants could not come to a consensus between the verbs "willing" and "inclined" in the *attitudes* item. In both cases, participants indicated that they felt compelled to answer high on the *attitudes* item and that although clinicians may be willing or inclined to use evidence, they may not actually do so in practice. Participants reported that the adjectives "willing" and "inclined" did not have equivalent translations in common French that would be suitable for a self-report measure (*enclin à* or *disposé à* are not commonly used words). Thus, the *attitudes* item was reworded from "I am willing to use EBP" to the notion of "it is worth the effort to [use EBP]". Appendix II reports the step-by-step changes at each step of the qualitative rewriting process.

Item modifications from the cognitive interview process

24 individual cognitive interviews were conducted with 10 PTs and 14 OTs in Canada (13 native English speakers, 12 native French speakers; one bilingual participant provided feedback in both languages). Appendix III presents an overview of the item evolution process during cognitive interviews. An overview of the modifications made to the items are described below.

The *self-efficacy* item underwent three iterations. From the initial item, "I am (very confident/somewhat/not very confident) in my ability to integrate evidence into my intervention plan", the word "integrate" was replaced with "apply" to be more action oriented. The words "intervention plan" were first replaced with "clinical cases" to avoid discriminating clinicians who solely perform assessments. The words "clinical cases" were then simplified to "practice" to avoid any confusion associated with the variability in clinical cases. Finally, the response option label "very confident" was changed to "confident" because participants stated it was difficult to endorse being very confident with one's ability to apply research evidence to practice. In the final version, the wording of two response option labels ("somewhat" and "not very" confident) was not an exact translation in French (*moyennement* and *peu confiant.e*).

The item on *use of research evidence* underwent three iterations. At the start of the cognitive interviews, this item consisted of asking respondents about the source of information, between research evidence, colleagues, or clinical experience, that they would seek when faced with a practice uncertainty. Participants found this item particularly difficult to answer because it was dependent on the case at hand (e.g., the availability of evidence for a clinical diagnosis or patient values) and the organizational context (e.g., whether colleagues were available and/or had experience related to the case). Clinicians reported that they often used a combination of all three sources and that it was difficult to select one to describe their typical behavior. The final wording of this item focused on the frequency of using research evidence when faced with a practice uncertainty. As a result, this modification avoids the conflicting response options of colleagues and clinical experience. Finally, an asterisk was added to define practice uncertainty as "a situation in which there is a gap in your knowledge relating to a clinical decision".

With six versions, the *attitudes* item underwent the highest number of iterations. At the start of the cognitive interviews, respondents were asked the extent to which incorporating evidence into practice was worth the effort. Participants suggested that the item not contain the connotation of "worth the effort" because it was (1) prone to social desirability bias (e.g., participants felt pressured to respond the best and highest level) and (2) did not translate well in French (e.g., *cela vaut l'effort* or *cela vaut la peine*). For these reasons, the item was modified to focus on the idea of EBP requiring effort (e.g., "It requires little/some/a lot of effort to integrate research evidence into practice"). The response option label "some (effort)" was changed to "moderate (effort)" to clarify the middle level response, and the words "(requires… effort) for me" were added to clarify the intent of eliciting the individual's perception of effort rather than a general belief. In French, the direct translation of "it requires little effort for me to…" is *cela me requiert peu d'efforts pour*… which was problematic for two reasons. First, starting a sentence in French with *cela* was too informal. Second, the verb *requiert* was too formal. The structure of the French sentence was changed to place the object (*intégrer les données probantes dans ma pratique*) before the verb and to replace *me requiert* with *me demande*.

The *resources* item underwent three iterations. The initial item was "I feel that I have the/only some of/do not have the necessary resources to integrate research evidence into my practice". Participants suggested omitting the words "I feel that …" and questioned which resources the item was referring to. An asterisk was added to clarify meaning and enumerate examples of resources facilitating EBP. In French, participants prefered the verb *je possède (les ressources nécessaires)* to *j'ai (les ressources nécessaires)*. For one response option label ("some of"), the final French wording was not an exact translation (*une partie des*).

Finally, the *activities* item underwent the least number of modifications. The only modification consisted of changing the words "consult research evidence" to "keep up to date with research evidence". Participants were interpreting the initial item as the frequency of using evidence in their practice which was already reflected in the *use of research evidence* item. The revised item reflects the concept of staying up to date with research evidence as an activity outside of routine CDM. Though three of the 24 participants suggested that we explicitly describe and quantify the three adverbs (regularly, occasionally, and rarely), the research team decided to avoid quantifying these adverbs as there is no agreed upon best practice for behavioral frequency of consulting the literature. By providing these three response options without specifying the exact range, the research team intended to capture clinician's self-report relative to their temporal understanding of keeping up to date with research evidence in their field. In French, "keeping up to date with research evidence" did not exactly translate, so the following modification was retained for conceptual equivalence *se tenir à jour quant aux données probantes*.

The initial instructional prompt was "For each group of statements, select ONE statement which best applies to you. Please respond as honestly as possible". The prompt was modified three times until the final version, "Please select ONE statement from each box which best reflects your current practice and context."

The visual presentation of the measure was improved following participant suggestions. Specifically, the lettering of each response option was bolded to make discriminating between levels easier. It was also suggested to number the five items (1 to 5) and letter the three response options (a, b, c) to reduce cognitive burden involved in completing the index. The final version of the PIRE-CDMI in English and French can be found in Appendix IV.

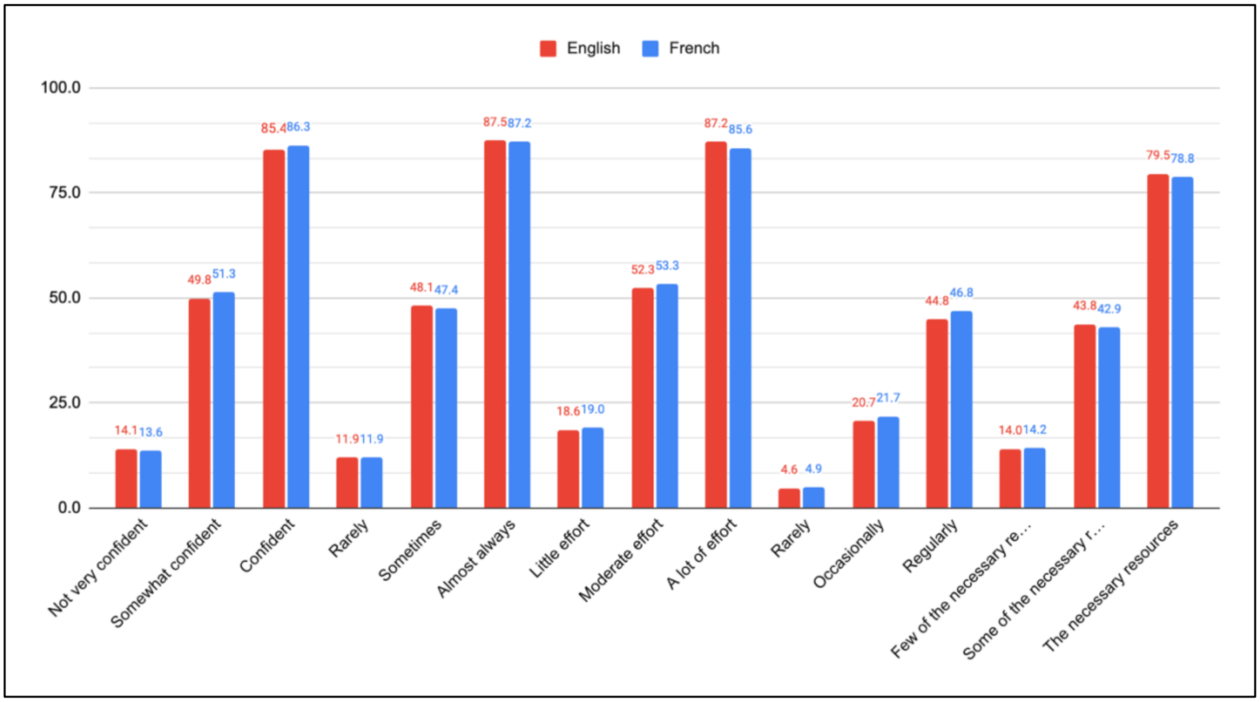
Scaling of response option labels

Among the 129 individuals who started the online survey, 60 were Canadian French native (46%) and 69 were English native (54%). Of the 129, 42 Francophones (32%) and 38 Anglophones (30%) were included for analysis. The rest were excluded due to incomplete surveys (n=25, 19%) and task miscomprehension (n=24, 19%). Descriptive results for the rating of the five response option sets by the 80 respondents is presented in Table 2. The ordinal nature of the ratings of response sets is illustrated in Figure 3. Multiple linear regression results, which are presented in Appendix V, did not suggest any important main effects of language on score for the five items, nor any important interaction of language and response option. Sparring the presence of a few outliers in all items, the residuals were normally distributed.

Table 2: Descriptive statistics for the rating of response options on a 0-100 scale (n=80)

	Mean (SD)	Min	Max	Mean (SD) in English (n=38)	Mean (SD) in French (n=42)
Item 1: I am _____ in my ability to apply research evidence to practice. **Anchors**: No confidence / Full confidence					
Not very confident	13.8 (5.2)	0	28	14.1 (3.7)	13.6 (6.2)
Somewhat confident	50.6 (6.6)	30	74	49.8 (6.3)	51.3 (7)
Confident	85.9 (5.9)	70	100	85.4 (4.7)	86.3 (6.9)
Item 2: When faced with a practice uncertainty, I ____ use research evidence. **Anchors**: None of the time / All of the time					
Rarely	11.9 (4.9)	0	26	11.9 (3.5)	11.9 (5.9)
Sometimes	47.7 (7.9)	25	75	48.1 (7.1)	47.4 (8.7)
Almost always	87.3 (5.3)	70	100	87.5 (4.1)	87.2 (6.3)
Item 3: It requires _____for me to integrate research evidence into practice. **Anchors**: No effort / Full effort					
Little effort	18.8 (5)	0	27	18.6 (4.5)	19 (5.4)
Moderate effort	52.9 (5.7)	38	70	52.3 (3.5)	53.3 (7.1)
A lot of effort	86.4 (4.8)	75	100	87.2 (4.1)	85.6 (5.2)
Item 4: I _____ keep up to date with research evidence. **Anchors**: 0 days/month and 30 days/month					
Rarely	4.8 (4.9)	0	17	4.6 (4.4)	4.9 (5.3)
Occasionally	21.3 (11.3)	3	60	20.7 (7.8)	21.7 (13.7)
Regularly	45.9 (18.2)	13	100	44.8 (16.2)	46.8 (20)
Item 5: I have _____ to integrate research evidence into my practice. **Anchors**: None / All imaginable resources					
Few of the necessary resources	14.1 (4.6)	3	25	14.0 (3.8)	14.2 (5.2)
Some of the necessary resources	43.4 (6)	20.0	56.0	43.8 (6.5)	44 (9.3)
The necessary resources	79.2 (7.2)	60.0	100.0	79.5 (7.2)	79.3 (8)
SD, standard deviation					

Figure 3. Histogram illustrating the frequency distribution of mean ratings of response option labels on a 0-100 scale for the five PIRE-CDMI items in English and French



Discussion

This study describes the item revision and rewriting process of a brief multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into CDM. Overall, the PIRE-CDMI underwent 12 iterations of qualitative rewriting. In the first phase, OTs, PTs and expert researchers rewrote the five items during an online focus group. The items were modified into declarative statements and changes were made to simplify the wording, increase relevance, and ensure coherence. Item rewriting was conducted simultaneously in English and French.

In the second phase, 24 cognitive interviews were conducted with practicing rehabilitation clinicians whereby the items underwent further revising to mitigate measurement error. Efforts were made to enhance the relevance of items to reflect the clinical reality of rehabilitation clinicians and to ensure that items reflected processes which were clear and actionable. Discussions with targeted end-users of the index allowed for the identification and resolution of problematic items at an ideal timing during the development process: after initial item development but prior to weight elicitation of dimensions and field testing[39,45]. Our sample size of 24 is in the high end of typical sample sizes for cognitive interviewing which typically range between 5 to 25 participants[36,39,40,49]. The interviewer followed the cognitive interview protocol and employed the verbal probing method which allowed the conversation to be shaped around specific topics and minimized the risk of participants digressing[39,42]. This approach facilitates the process of reviewing items by placing less cognitive burden on the participants who are guided through the discussion.

As a result of the steps described in this paper, important changes were made to all items included in the PIRE-CDMI. The *use of research evidence* item changed considerably throughout the process and iterations included probes about (1) the frequency of integrating the three EBP pillars (research evidence, clinical expertise, and patient preferences) and (2) the primary and most reliable source of knowledge (research evidence, colleagues, and clinical experience). These two iterations were very difficult for participants to answer and failed to produce useful information. It is not surprising that clinicians attested to integrating all three pillars of EBP into CDM and relying on all proposed sources of knowledge to various extents. In fact, the tripartite definition is foundational to how rehabilitation clinicians conceptualize EBP[50]. However, asking clinicians to select their most relied upon pillar of EBP or to determine the frequency at which they integrate the three components is anathema to the reality of CDM whereby these elements are inextricably intertwined[23]. Indeed, Thornton[17] posits that the three elements of EBP form a unifiable whole despite the model illustrating them as being disparate elements.

Asking clinicians to select the most relied-upon source of knowledge (with one of the response options being research evidence) in a measure relating to EBP appears to introduce high levels of social desirability bias, a type of bias due to respondents' desire to be perceived in a favorable regard to others[51,52]. For example, a respondent may interpret the desirable answer to be "research evidence", given the title of the index, and the undesirable answer to be "colleague"' or "clinical experience"' through the natural process of deduction. This item formulation could inadvertently imply that consulting a colleague or relying on clinical experience is ill-advised when, in fact, these sources of knowledge are foundational to being a competent, reflexive, and evidence-based clinician[17,24,50,53].

Instead, the final version of this item focused on the frequency of using research evidence when faced with a clinical uncertainty. Removing item probes relating to clinical experience, colleagues or patient preferences should not be interpreted as a disregard to these vital sources of knowledge. Indeed, I acknowledge and value the plurality of knowledge in CDM or the multiple “ways of knowing that have the potential to inform the scope of professional activities”[54(p255)]. Yet, rehabilitation clinicians continue to voice challenges related to the integration of research evidence into CDM signaling a deep-rooted need for support with this component of EBP[21,24,26,27].

The idea of a “clinical uncertainty” was introduced to contextualize the behavior of seeking and using research evidence in CDM. The addition of contextual cues in items has been said to increase the validity of responses, notably when the behaviour has an element of automaticity[55]. CDM often relies on automatic and intuitive reasoning rather than analytical reasoning, a phenomenon which is hypothesized to become stronger over time[56,57]. To tap into clinicians’ analytical reasoning, we contextualized the item with a trigger for reflection or a call to action (a clinical uncertainty) to capture clinicians’ behavioral tendency to seek out and consider research evidence. This clinical uncertainty can be compared to *the event* proposed in the reflective practice literature which is defined as “an event that occurs in everyday practice […] that leaves the occupational therapist with the urge to revisit it to make sense of it for the benefit of his or her future practice."[58(p345)]. The need to nuance this item with a clinical uncertainty is further reinforced by a possible mechanism whereby, over time, research-based knowledge becomes consolidated into tacit or experiential knowledge; in such cases, it may no longer be distinguished as *research evidence* but rather transformed into expert practice that is adapted to the practice context[59,60]. Correspondingly, it may be difficult for clinicians to discern how frequently they use research evidence on the day to day. Lastly, consulting research evidence in everyday practice may not be a desirable behavior as it could conceal other professional difficulties such as low confidence in one’s clinical reasoning abilities[61]. Thus, the more compelling question is not so much whether clinicians consult formal sources of evidence every day, but whether they do so when confronted with a gap in their knowledge.

Through the course of the interviews, many modifications were made to the *attitudes* item to remedy the social desirability bias that was reported by interview participants. Despite changes to the item, most interview participants continued to report that EBP was a desirable process; they felt compelled to select the highest response option for attitudes, especially given that their professional entry-level training strongly emphasized EBP. Attitudes towards EBP have been defined as “the values ascribed by the [clinician] to the importance and usefulness of EBP to inform clinical decision-making”[62(p4)]. The evidence demonstrating the relationship between attitudes towards EBP and EBP behavior is inconclusive. While some studies have suggested that holding positive attitudes towards EBP is an important precursor to EBP behavior [63–67], others have demonstrated that they do not translate into effective EBP behavior[20–22,68–71]. Although exploring the reasons why this evidence is inconclusive is beyond the scope of this paper, we postulate that measuring attitudes towards EBP may not be useful in the context of this brief multidimensional index given that (1) the relationship between attitudes and EBP behaviors is uncertain; (2) it is well-established that rehabilitation clinicians are generally convinced of the value of EBP and believe it to be a desirable and necessary process[20–22,69,71–73]; and (3) value-laden items which can prejudice respondents should be omitted from measures[36] and attitudes are inherently value-laden. Furthermore, when assessing attitudes for predicting behaviors, it is recommended to avoid measuring attitudes towards a general concept and to focus on specific behaviors[74]. For these reasons, the focus of this item was changed from the perceived importance of EBP towards the required effort to pursue EBP, an item formulation that participants did not flag as being prone to social desirability bias. Using *effort* instead of *attitudes* circumvents asking clinicians whether they consider EBP to be valuable (which clinicians commonly affirm) and highlights the perceived cost of integrating research evidence into practice[75]. *Effort* is defined by the Cambridge Dictionary as "physical or mental activity needed to achieve something". Social psychology and behavioral theorists have identified effort as largely contributing to behavioral motivation[76–78]. People are less likely to engage in a behavior if it requires a large amount of effort[79] and it is clear that rehabilitation clinicians perceive the enactment of EBP as being effortful (e.g., having to put in the time to read the literature despite prevalent barriers and increasing professional demands on clinicians)[23,24,69]. Moreover, a feeling of enjoyment, worth and resource replenishment can confound the perception of effort[79]. If clinicians value the outcome of EBP (e.g., positive patient outcomes, increased professional credibility), it may require less of an effort to enact EBP. Consequently, using effort may act as a proxy for attitude or perceived value without the social desirability bias.

The *activities* item changed from the initial form, "In the past month, how often have you: made time to read research?" to "I regularly/occasionally/rarely keep up to date with research evidence." As focus group participants stated, it was important that this item remain open to various sources of research evidence outside of scientific articles. As such, one could associate keeping up to date with research evidence to leading or assisting a journal club, reading email subscription alerts, or gaining research-based knowledge from a colleague. This departure from formal sources of research-based knowledge is more aligned with how rehabilitation clinicians typically gather research evidence. Indeed, OTs and PTs favor informal, quick methods of gaining research evidence, and tend to keep up with research evidence through a variety of informal sources including consultation with trusted peers and email reminders[15,23,24,47,80–82]. The process by which rehabilitation clinicians rely on colleagues for research-based knowledge is starting to gain importance in the EBP literature as a recognized and beneficial mechanism of EBP[24,82–84].

In light of the rapidly evolving knowledge base in rehabilitation research[85], we posit that the idea of staying up to date with research evidence is vital to being an evidence-based practitioner. As knowledge is produced at different rates for different areas of practice, and that this index can be used with clinicians from various settings, no explicit frequency denominator was attributed to this item. For instance, "regularly" could mean once every two months for a clinician in stroke rehabilitation or once a year for a clinician in palliative care. Our intent with this item is to capture the respondent's self-rating relative to their understanding of what "regularly, occasionally and rarely" mean relative to their field of practice and relative to their perception of what is feasible given their clinical reality.

The *self-efficacy* item initially asked for respondents' confidence "in [their] current level of ability to determine if the evidence from the research literature applies to [their] patient?". This item was modified to "ability to apply research evidence to practice" to reflect a slightly broader skill of interest, one that has more relevance to clinicians. This change mirrors contemporary views on how EBP is enacted in clinical practice which is less restrictive to the discrete technical skill of critical appraisal and more about how research evidence applies to practice[80,83,86–89]. Self-efficacy has been identified as one of the key constructs in influencing EBP behavior[90,91]. Applied to the context of EBP, *self-efficacy theory* developed by Bandura[92,93] would suggest that clinicians' judgment of their ability to engage in the EBP process is a major influence on their decisions to embrace or avoid the integration of research evidence into CDM[91,94]. As such, if clinicians' confidence in their abilities increases, their propensity to integrate research evidence into CDM is expected to increase. Our results highlight that clinicians have difficulty endorsing the "very confident" and "not confident" response options because they were perceived as being non desirable traits. For this reason, these extreme qualifiers were replaced with more modest qualifiers (“confident” and “not very confident”). This important finding must be considered when interpreting information from other scales with these extreme qualifiers as respondents may tend to endorse the middle response option.

The *resources* item was modified from “my organization (supports best practice)” to “I have/have some of/have few of the necessary resources to integrate research evidence into my practice”. The initial item probed respondents about their organization and was inherently prone to social desirability bias due to the inherent power differentials present in an employer-employee dynamic. Respondents may have felt intimated to respond positively towards their organization. We reduced the emphasis on the organization by omitting the word and shifting the focus towards having adequate affordances (resources) to be evidence-based. An asterisk was added to provide examples of resources (e.g., paid time, access to a computer and access to necessary therapeutic material). The intent with this item is to capture a clinician’s perception of the external affordances of EBP as it is well-known that the environment, specifically the availability of organizational resources, plays an important role in the acquisition and application of research-evidence[15,21,26,27,50,71,95–97]. We acknowledge that the organizational culture and leadership also largely influence propensity to integrate research evidence into CDM, however we hypothesize that such an item would also be prone to social desirability bias. Responding to an item on available resources may be perceived as being more constructive and can still be an indicator of the culture and leadership.

Based on our survey findings, the effect of language on VAS ratings was trivial, which supports the equivalency of the response option labels in English and French for all PIRE-CDMI items. The distribution of ratings also demonstrates the ordinal consistency of response options and the quasi-interval nature of the scales (i.e., response options are equally spaced). The largest difference in ratings between English and French translations was for the response option ‘‘regularly” (*keeping up to date* item), which was rated two points higher in French than in English category. Despite this being the largest difference, a two-point difference on 100 is considered trivial. Other studies using the same response scaling method considered a difference of over five points on the 0–100 VAS to be indicative of requiring possible revision[98]. Interestingly, the *keeping up to date* item also had the largest variation in ratings within the same language, especially for the “occasionally” and “regularly” response options. This response set required respondents to rate adverbs denoting frequency (regularly, occasionally, rarely) on a scale from 0 days/month to 30 days/month. A possible explanation for this heightened variability in ratings may be because of the subjective nature of these behavioral frequency adverbs (potentially dependent on the area of practice) and lack of consensus on how often rehabilitation clinicians should keep up to date with the literature. Our survey presents preliminary information on how these adverbs may be quantified by future and current healthcare professionals with respect to keeping up to date with research evidence; “regularly”, “occasionally” and “rarely” were rated, on average, 14, 6 and 1 day(s)/month, respectively. However, these findings must be interpreted with caution since no demographic information was gathered regarding profession or area of practice.

A strength of this study is the multi-phased rigorous qualitative review process which included target end-users and allowed for the development of relevant and clear items. The recruitment strategy for the response option scaling survey provided us with a reasonable sample size within a month. In developing the survey, we aimed to provide adequate guidance to maximize respondents' comprehension of the task. Pilot testing enabled us to add examples and clarify the instructions. Still, given that 25 individuals did not complete the survey (19%) and that 24 individuals had to be excluded due to apparent miscomprehension of the task (19%), this exercise may have been perceived as difficult and burdensome, a finding also reported by others in the context of a valuation exercise for the EQ-5D[99]. The task involved an unfamiliar method of placing response option labels on a 0 to 100 scale which required an ability for abstract reasoning. This speaks to the acceptability of response scaling for a target population of healthcare professionals. Given the lack of available demographic data, it is impossible to discern who misunderstood the task. Comprehension may have been improved with a quick instructional video.

Before deploying the PIRE-CDMI, there remains an important developmental step which consists of estimating the relative weights of each dimension-level. This will allow for the generation of a more accurate total score that takes into consideration end-users’ perceived relative contribution of dimensions on the overall construct of propensity to integrate research evidence into CDM. Despite our best efforts at mitigating systematic measurement error during this item review phase, future research aimed at field testing the PIRE-CDMI with various samples of OTs and PTs (e.g., different clinical settings) is required before inferences can be drawn in clinical contexts. We acknowledge that the initial mathematical properties of the prototype PIRE-CDMI established in chapter 5 may have changed due to item rewriting. While this must be confirmed in future testing, our findings pertaining to the quasi-interval spacing of response option labels gives us reason to believe that the interval properties of the scale still hold. Due to important linguistic differences between different countries, we suggest undergoing a thorough cultural adaptation and reassessment before using the PIRE-CDMI with English and French-speaking individuals outside of Canada.

Finally, while some authors have stated that short scales are a limitation and can compromise the validity and reliability of inferences drawn from a measure[100,101], others have found value in the efficiency of short scales[102,103]. There exists a delicate trade-off between scale comprehensiveness and feasibility. Given the resource-strained healthcare context, the aim was to create a short index capable of rapidly estimating elements of EBP requiring improvement. This measure can be used as an efficient global outcome measure of a clinician's propensity to integrate research evidence into CDM for research purposes and professional self-reflection which may then be complemented with more comprehensive and lengthier measures. Intervention strategies can then be developed to target the specific areas requiring support.

Conclusion

The three consecutive phases described in this paper illustrate a rigorous approach to developing a brief multidimensional index of propensity to integrate research evidence into CDM in rehabilitation that is coherent, clear, and relevant to Canadian OTs and PTs. A focus group and cognitive interviews with end-users gave rise to important item modifications in English and French to minimize ambiguity, measurement bias and cognitive burden on respondents. Finally, response option labels in English and French were found to be equivalent through a cross-sectional online survey wherein response option labels were compared on 0 to 100 scales in both languages. The next developmental step involves estimating the relative weights of dimensions to generate a scoring algorithm for the index.

References

1. CAOT. Profile of occupational therapist practice in Canada. Published online 2012. Accessed September 17, 2020. https://www.caot.ca/document/3653/2012otprofile.pdf

2. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1

3. Emparanza JI, Cabello JB, Burls AJE. Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *J Eval Clin Pract*. 2015;21(6):1059-1065. doi:10.1111/jep.12460

4. Group NPA. Competency profile for physiotherapists in Canada. *Natl Physiother Advis Group*. Published online 2017. https://www.peac-aepc.ca/pdfs/Resources/Competency%20Profiles/Competency%20Profile%20for%20PTs%202017%20EN.pdf

5. Haynes RB, Sackett DL, Gray JMA, Cook DJ, Guyatt GH. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. *Evid Based Med*. 1996;1(7):196.

6. Haynes RB, Sackett DL, Richardson WS, Rosenberg W, Langley GR. Evidence-based medicine: How to practice & teach EBM. *Can Med Assoc J*. 1997;157(6):788.

7. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn’t. *Br Med J*. Published online 1996:71-72.

8. Sackett DL, Wennberg JE. *Choosing the Best Research Design for Each Question: It’s Time to Stop Squabbling over the “Best” Methods*. British Medical Journal Publishing Group; 1997.

9. DiCenso A, Ciliska D, Guyatt G. Introduction to evidence-based nursing. *Evid-Based Nurs Guide Clin Pract*. Published online 2005:3-19.

10. Gutenbrunner C, Nugraha B. Decision-Making in Evidence-Based Practice in Rehabilitation Medicine: Proposing a Fourth Factor. *Am J Phys Med Rehabil*. 2020;99(5):436-440. doi:10.1097/PHM.0000000000001394

11. Kitson A, Harvey G, McCormack B. Enabling the implementation of evidence based practice: a conceptual framework. *BMJ Qual Saf*. 1998;7(3):149-158.

12. Satterfield JM, Spring B, Brownson RC, et al. Toward a Transdisciplinary Model of Evidence-Based Practice. *Milbank Q*. 2009;87(2):368-390. doi:10.1111/j.1468-0009.2009.00561.x

13. Craik J, Rappolt S. Theory of Research Utilization Enhancement: A Model for Occupational Therapy. *Can J Occup Ther*. 2003;70(5):266-275. doi:10.1177/000841740307000503

14. Craik J, Rappolt S. Enhancing research utilization capacity through multifaceted professional development. *Am J Occup Ther*. 2006;60(2):155-164.

15. Robertson L, Graham F, Anderson J. What Actually Informs Practice: Occupational Therapists' Views of Evidence. *Br J Occup Ther*. 2014;76:317-324. doi:10.4276/030802213X13729279114979

16. Kothari AR, Bickford JJ, Edwards N, Dobbins MJ, Meyer M. Uncovering Tacit Knowledge: A Pilot Study to Broaden the Concept of Knowledge in Knowledge Translation. *BMC Health Serv Res*. 2011;11(1):1-10. doi:10.1186/1472-6963-11-198

17. Thornton T. Tacit knowledge as the unifying factor in evidence based medicine and clinical judgement. *Philos Ethics Humanit Med*. 2006;1(1):2. doi:10.1186/1747-5341-1-2

18. Kashaf MS, McGill E. Does shared decision making in cancer treatment improve quality of life? A systematic literature review. *Med Decis Making*. 2015;35(8):1037-1048.

19. Shay LA, Lafata JE. Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes. *Med Decis Making*. 2015;35(1):114-131. doi:10.1177/0272989X14551638

20. Bernhardsson S, Johansson K, Nilsen P, Öberg B, Larsson MEH. Determinants of Guideline Use in Primary Care Physical Therapy: A Cross-Sectional Survey of Attitudes, Knowledge, and Behavior. *Phys Ther*. 2014;94(3):343-354. doi:10.2522/ptj.20130147

21. da Silva T, da Cunha Menezes Costa L, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. 2015;20(3):388-401. doi:10.1016/j.math.2014.10.009

22. Graham F, Robertson L, Anderson J. N ew Z ealand occupational therapists' views on evidence-based practice: A replicated survey of attitudes, confidence and behaviours. *Aust Occup Ther J*. 2013;60(2):120-128.

23. Jeffery H, Robertson L, Reay K. Sources of evidence for professional decision-making in novice occupational therapy practitioners: clinicians' perspectives. *Br J Occup Ther*. 2020;84:030802262094139. doi:10.1177/0308022620941390

24. Rochette A, Brousseau M, Vachon B, Engels C, Amari F, Thomas A. What occupational therapists' say about their competencies' enactment, maintenance and development in practice? A two-phase mixed methods study. *BMC Med Educ*. 2020;20. doi:10.1186/s12909-020-02087-4

25. Upton D, Stephens D, Williams B, Scurlock-Evans L. Occupational Therapists' Attitudes, Knowledge, and Implementation of Evidence-Based Practice: A Systematic Review of Published Research. *Br J Occup Ther*. 2014;77:24. doi:10.4276/030802214X13887685335544

26. Juckett LA, Wengerd LR, Faieta J, Griffin CE. Evidence-Based Practice Implementation in Stroke Rehabilitation: A Scoping Review of Barriers and Facilitators. *Am J Occup Ther Off Publ Am Occup Ther Assoc*. 2020;74(1):7401205050p1-7401205050p14. doi:10.5014/ajot.2020.035485

27. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to Evidence-Based Practice Implementation in Physiotherapy: A Systematic Review and Meta-Analysis. *Int J Qual Health Care*. 2021;(mzab093). doi:10.1093/intqhc/mzab093

28. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gómez JE. Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the literature: Evidence-based physiotherapy instruments. *J Eval Clin Pract*. 2014;20(6):767-778. doi:10.1111/jep.12180

29. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci*. 2014;9(1):118. doi:10.1186/s13012-014-0118-8

30. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

31. Buchanan H, Siegfried N, Jelsma J. Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review: Evidence-based Practice Survey Instruments. *Occup Ther Int*. 2016;23(2):59-90. doi:10.1002/oti.1398

32. Glegg SMN, Holsti L. Measures of knowledge and skills for evidence-based practice: A systematic review. *Can J Occup Ther Ott*. 2010;77(4):219-232.

33. Roberge-Dao J, Maggio LA, Zaccagnini M, et al. Quality, methods, and recommendations of systematic reviews on measures of evidence-based practice: an umbrella review. *JBI Evid Synth*. 2022;20(4):1004-1073.

34. McGorry SY. Measurement in a cross-cultural environment: survey translation issues. *Qual Mark Res Int J*. Published online 2000.

35. Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychol Assess*. 1995;7(3):238.

36. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, USA; 2015.

37. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417-1432.

38. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing*.; 2014.

39. Willis GB. *Analysis of the Cognitive Interview in Questionnaire Design*. Oxford University Press; 2015.

40. Beatty PC, Willis GB. Research synthesis: The practice of cognitive interviewing. *Public Opin Q*. 2007;71(2):287-311.

41. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health*. 2018;6. doi:10.3389/fpubh.2018.00149

42. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*. 2003;12(3):229-238.

43. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *J Adv Nurs*. 2003;42(1):57-63.

44. Kuspinar A, Bouchard V, Moriello C, Mayo NE. Development of a Bilingual MS-Specific Health Classification System. *Int J MS Care*. 2016;18(2):63-70. doi:10.7224/1537-2073.2014-106

45. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. Published online 2016.

46. Emmel N. *Sampling and Choosing Cases in Qualitative Research: A Realist Approach*. SAGE; 2013.

47. Keller SD, Ware JE, Gandek B, et al. Testing the Equivalence of Translations of Widely Used Response Choice Labels: Results from the IQOLA Project. Published online 1998:12.

48. Luo N, Li M, Chevalier J, Lloyd A, Herdman M. A comparison of the scaling properties of the English, Spanish, French, and Chinese EQ-5D descriptive systems. *Qual Life Res*. 2013;22(8):2237-2243. doi:10.1007/s11136-012-0342-0

49. Sheatsley PB. Questionnaire construction and item writing. *Handb Surv Res*. 1983;4(1):195-230.

50. Hallé MC, Mylopoulos M, Rochette A, et al. Attributes of evidence-based occupational therapists in stroke rehabilitation. *Can J Occup Ther*. 2018;85(5):351-364. doi:10.1177/0008417418802600

51. Fisher RJ. Social desirability bias and the validity of indirect questioning. *J Consum Res*. 1993;20(2):303-315.

52. King MF, Bruner GC. Social desirability bias: A neglected aspect of validity testing. *Psychol Mark*. 2000;17(2):79-103.

53. Buetow S, Kenealy T. Evidence-based medicine: the need for a new definition: Examining the meaning of EBM. *J Eval Clin Pract*. 2000;6(2):85-92. doi:10.1046/j.1365-2753.2000.00237.x

54. Kinsella EA, Whiteford GE. Knowledge generation and utilisation in occupational therapy: Towards epistemic reflexivity. *Aust Occup Ther J*. 2009;56(4):249-258.

55. Presseau J, McCleary N, Lorencatto F, Patey AM, Grimshaw JM, Francis JJ. Action, actor, context, target, time (AACTT): a framework for specifying behaviour. *Implement Sci*. 2019;14(1):102. doi:10.1186/s13012-019-0951-x

56. Potthoff S, Rasul O, Sniehotta FF, et al. The relationship between habit and healthcare professional behaviour in clinical practice: a systematic review and meta-analysis. *Health Psychol Rev*. 2019;13(1):73-90. doi:10.1080/17437199.2018.1547119

57. Reyna VF. A theory of medical decision making and health: fuzzy trace theory. *Med Decis Making*. 2008;28(6):850-865.

58. Bannigan K, Moores A. A model of professional thinking: Integrating reflective practice and evidence based practice. *Can J Occup Ther*. 2009;76(5):342-350.

59. Paterson M, Higgs J, Wilcox S. Developing expertise in judgement artistry in occupational therapy practice. *Br J Occup Ther*. 2006;69(3):115-123.

60. Paterson M, Wilcox S, Higgs J. Exploring dimensions of artistry in reflective practice. *Reflective Pract*. 2006;7(4):455-468.

61. Thomas A, Chin-Yee B, Mercuri M. Thirty years of teaching evidence-based medicine: have we been getting it all wrong? *Adv Health Sci Educ*. 2022;27(1):263-276.

62. Tilson JK, Kaplan SL, Harris JL, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Educ*. 2011;11(1):1-10. doi:10.1186/1472-6920-11-78

63. Aarons GA. Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Ment Health Serv Res*. 2004;6(2):61-74.

64. Ajzen I. From intentions to actions: A theory of planned behavior. In: *Action Control*. Springer; 1985:11-39.

65. Heiwe S, Kajermo KN, Tyni-Lenné R, et al. Evidence-based practice: attitudes, knowledge and behaviour among allied health care professionals. *Int J Qual Health Care*. 2011;23(2):198-209. doi:10.1093/intqhc/mzq083

66. Melas CD, Zampetakis LA, Dimopoulou A, Moustakis V. Evaluating the properties of the Evidence-Based Practice Attitude Scale (EBPAS) in health care. *Psychol Assess*. 2012;24(4):867.

67. Rye M, Torres EM, Friborg O, Skre I, Aarons GA. The Evidence-based Practice Attitude Scale-36 (EBPAS-36): a brief and pragmatic measure of attitudes to evidence-based practice validated in US and Norwegian samples. *Implement Sci*. 2017;12(1):44.

68. Arumugam V, MacDermid JC, Walton D, Grewal R. Attitudes, knowledge and behaviors related to evidence-based practice in health professionals involved in pain management. *JBI Evid Implement*. 2018;16(2):107-118. doi:10.1097/XEB.0000000000000131

69. Hitch D, Nicola-Richmond K, Richards K, Stefaniak R. Student perspectives on factors that influence the implementation of evidence-based practice in occupational therapy. *JBI Evid Implement*. 2021;19(4):409-418. doi:10.1097/XEB.0000000000000285

70. Iles R, Davidson M. Evidence based practice: a survey of physiotherapists' current practice. *Physiother Res Int*. 2006;11(2):93-103.

71. Scurlock-Evans L, Upton P, Upton D. Evidence-Based Practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014;100(3):208-219. doi:10.1016/j.physio.2014.03.001

72. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014

73. Thomas A, Zoubi FA, Mayo NE, et al. Individual and organizational factors associated with evidence-based practice among physical and occupational therapy recent graduates: A cross-sectional national study. *J Eval Clin Pract*. 2020;n/a(n/a). doi:https://doi.org/10.1111/jep.13518

74. Fishman J, Yang C, Mandell D. Attitude theory and measurement in implementation science: a secondary review of empirical studies and opportunities for advancement. *Implement Sci*. 2021;16(1):87. doi:10.1186/s13012-021-01153-9

75. Eccles J. Expectancies, values and academic behaviors. *Achiev Achiev Motiv*. Published online 1983.

76. Contreras-Huerta LS, Pisauro MA, Apps MAJ. Effort shapes social cognition and behaviour: A neuro-cognitive framework. *Neurosci Biobehav Rev*. 2020;118:426-439. doi:10.1016/j.neubiorev.2020.08.003

77. Croxson PL, Walton ME, O'Reilly JX, Behrens TE, Rushworth MF. Effort-based cost–benefit valuation and the human brain. *J Neurosci*. 2009;29(14):4531-4541.

78. Kurzban R, Duckworth A, Kable JW, Myers J. An opportunity cost model of subjective effort and task performance. *Behav Brain Sci*. 2013;36(6):661-679.

79. Inzlicht M, Shenhav A, Olivola CY. The Effort Paradox: Effort Is Both Costly and Valued. *Trends Cogn Sci*. 2018;22(4):337-349. doi:10.1016/j.tics.2018.01.007

80. Condon C, McGrane N, Mockler D, Stokes E. Ability of physiotherapists to undertake evidence-based practice steps: a scoping review. *Physiotherapy*. 2016;102(1):10-19.

81. Myers CT, Lotz J. Practitioner Training for Use of Evidence-Based Practice in Occupational Therapy. *Occup Ther Health Care*. 2017;31(3):214-237. doi:10.1080/07380577.2017.1333183

82. Iqbal MZ, Rochette A, Mayo NE, et al. Exploring how evidence-based practice of occupational and physical therapists evolves: A longitudinal mixed methods national study. *Submitted to Plos One on July 08, 2022 (PONE-D-22-19314).*

83. Gabbay J, May A le. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. *BMJ*. 2004;329(7473):1013. doi:10.1136/bmj.329.7473.1013

84. Salter KL, Kothari A. Knowledge 'Translation'as social learning: negotiating the uptake of research-based knowledge in practice. *BMC Med Educ*. 2016;16(1):1-10.

85. Negrini S, Levack W, Gimigliano F, Arienti C, Villafañe JH, Kiekens C. The struggle for evidence in physical and rehabilitation medicine: publication rate of randomized controlled trials and systematic reviews is growing more than in other therapeutic fields. *Am J Phys Med Rehabil*. 2019;98(4):258-265.

86. Ghali WA, Saitz R, Sargious PM, Hershman WY. Evidence-based medicine and the real world: understanding the controversy. *J Eval Clin Pract*. 1999;5(2):133-138. doi:10.1046/j.1365-2753.1999.00183.x

87. Ghali WA, Sargious PM. The evolving paradigm of evidence-based medicine. *J Eval Clin Pract*. 2002;8(2):109-112. doi:10.1046/j.1365-2753.2002.00364.x

88. Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence based care. *BMJ*. 2000;320(7240):954-955.

89. Tikkinen KA, Guyatt GH. Understanding of research results, evidence summaries and their applicability—not critical appraisal—are core skills of medical curriculum. *BMJ Evid-Based Med*. Published online 2021.

90. Michie S, Johnston M, Abraham C, et al. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care*. 2005;14(1):26-33. doi:10.1136/qshc.2004.011155

91. Salbach NM, Jaglal SB %J J of evaluation in clinical practice. Creation and validation of the evidence-based practice confidence scale for health care professionals. 2011;17(4):794-800.

92. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*. 1977;84(2):191.

93. Bandura A, Wessels S. *Self-Efficacy*. Vol 4. na; 1994.

94. Blackman IR, Giles T. Psychometric Evaluation of a Self-Report Evidence-Based Practice Tool Using Rasch Analysis. *Worldviews Evid Based Nurs*. 2015;12(5):253-264. doi:10.1111/wvn.12105

95. Lyons C, Brown T, Tseng MH, Casey J, McDonald R. Evidence-based practice and research utilisation: Perceived research knowledge, attitudes, practices and barriers among Australian paediatric occupational therapists. *Aust Occup Ther J*. 2011;58(3):178-186.

96. McCluskey A, Vratsistas-Curto A, Schurr K. Barriers and enablers to implementing multiple stroke guideline recommendations: a qualitative study. *BMC Health Serv Res*. 2013;13(1):323. doi:10.1186/1472-6963-13-323

97. Salbach NM, Guilcher SJ, Jaglal SB, Davis DA. Determinants of research use in clinical decision making among physical therapists providing services post-stroke: a cross-sectional study. *Implement Sci*. 2010;5(1):77. doi:10.1186/1748-5908-5-77

98. Scott NW, Etta JA, Aaronson NK, et al. An evaluation of the response category translations of the EORTC QLQ-C30 questionnaire. *Qual Life Res*. 2013;22(6):1483-1490. doi:10.1007/s11136-012-0276-6

99. Devlin NJ, Hansen P, Selai C. Understanding health state valuations: a qualitative analysis of respondents' comments. *Qual Life Res*. 2004;13(7):1265-1277.

100. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. *Psicol Reflex E Crítica*. 2017;30(1):3. doi:10.1186/s41155-016-0057-1

101. Raykov T. Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *Br J Math Stat Psychol*. 2008;61(2):275-285.

102. Rammstedt B, Beierlein C. Can't we make it any shorter? *J Individ Differ*. Published online 2014.

103. Ziegler M, Kemper C, Kruyen P. Short Scales - Five Misunderstandings and Ways to Overcome Them. *J Individ Differ*. 2014;35:185-189. doi:10.1027/1614-0001/a000148

Appendices

Appendix I. Five PIRE-CDMI response sets

Dimension	Response options	Response anchors
Use of research evidence	Almost always Sometimes Rarely	None of the time / All of the time
Self-efficacy	Confident Somewhat confident Not very confident	No confidence / full confidence
Resources	Have the necessary resources Have some of the necessary resources Have few of the necessary resources	None / All imaginable resources
Attitudes → Effort	Little effort Moderate effort A lot of effort	No effort / Full effort
Activities related to EBP → Keeping up to date	Regularly Occasionally Rarely	0 days/month / 30 days/month

Appendix II. Step-by-step changes at each step of the item rewriting process in both languages for each item

English version of the index

Item: Self-reported use of research evidence				
Version 1	How often have you done each of the following activities in the past month?	Decide on an appropriate course of action based on integrating the research evidence, clinical judgment and patient or client preferences?	0 (Never) / 1 (One time or more)	
Focus group	Changed the initial item structure to three declarative statements. Changed "[integrating the 3 pillars]" to "integrate research evidence". Changed "decide on an appropriate course of action" to "into the care plan".			
Version 2	I regularly integrate research evidence into the care plan.	I occasionally integrate research evidence into the care plan.	I rarely integrate research evidence into the care plan.	
Research team	Changed item to elicit the first source of information when faced with a practice uncertainty.			
Version 3	When faced with a practice uncertainty, I rely on research evidence.	When faced with a practice uncertainty, I rely on my colleagues.	When faced with a practice uncertainty, I rely on my clinical experience.	
Cognitive interviews	Changed the sentence structure.			
Version 4	I first rely on **research evidence** when faced with a practice uncertainty.	I first rely on **my colleagues** when faced with a practice uncertainty.	I first rely on **my clinical experience** when faced with a practice uncertainty.	
Version 5	Idem			
Cognitive interviews	Added an asterisk to define what we mean by "practice uncertainty"			
Version 6	I first rely on **research evidence** when faced with a practice uncertainty.	I first rely on **my colleagues** when faced with a practice uncertainty.	I first rely on **my clinical experience** when faced with a practice uncertainty.	**Uncertainty: a situation in which there is a gap in your knowledge relating to a clinical decision*
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			

Cognitive interviews	To decrease variability due to clinical population and context, the item has been modified to focus on frequency of using research evidence when faced with a practice uncertainty.			
Version 11	When faced with a practice uncertainty*, a) I **almost always** use research evidence.	b) I **sometimes** use research evidence.	c) I **rarely** use research evidence.	**Uncertainty: a situation in which there is a gap in your knowledge relating to a clinical decision*
Version 12	Idem			
Item: Self-efficacy				
Version 1	Please indicate how confident you are in your current level of ability by choosing the corresponding number on the following rating scale.	Determine if the evidence from the research literature applies to your patient	0 (0-25%) / 1 (25-50%) / 2 (50-75%) /3 (75-100%)	
Focus group	Changed the initial item structure to three declarative statements. Omitted “in your current level of ability” to shorten the sentence. Changed “evidence from the research literature” to “research evidence” to be inclusive to other sources of research evidence.			
Version 2	I am confident in deciding if research evidence applies to my patient*.	I am somewhat confident in deciding if research evidence applies to my patient*.	I am not confident in deciding if research evidence applies to my patient*.	
Research team	Changed "deciding if research evidence applies to my patient" to "in my ability to integrate research evidence into my intervention plan". Added "very" superlative to first and last level.			
Version 3	I am very confident in my ability to integrate research evidence into my intervention plan.	I am somewhat confident in my ability to integrate research evidence into my intervention plan.	I am not very confident in my ability to integrate research evidence into my intervention plan.	
Version 4	I am **very confident** in my ability to integrate research evidence into my intervention plan.	I am **somewhat confident** in my ability to integrate research evidence into my intervention plan.	I am **not very confident** in my ability to integrate research evidence into my intervention plan.	
Cognitive interviews	Changed "integrate" for "apply". Changed 'intervention plan' for 'clinical cases			
Version 5	I am **very confident** in my ability to apply research evidence to clinical cases.	I am **somewhat confident** in my ability to apply research evidence to clinical cases.	I am **not very confident** in my ability to apply research evidence to clinical cases.	
Version 6	Idem			

Version 7	Idem			
Version 8	Idem			
Cognitive interviews	Changed 'very confident" to "confident"			
Version 9	I am **confident** in my ability to apply research evidence to clinical cases.	I am **somewhat confident** in my ability to apply research evidence to clinical cases.	I am **not very confident** in my ability to apply research evidence to clinical cases.	
Cognitive interviews	Changed "clinical cases" for "practice"			
Version 10	I am **confident** in my ability to apply research evidence to practice.	I am **somewhat confident** in my ability to apply research evidence to practice.	I am **not very confident** in my ability to apply research evidence to practice.	
Version 11	Idem			
Version 12	Idem			
Item: Resources				
Version 1	Please indicate your level of agreement with the following statements:	My organization supports best practice	0 (Strongly disagree /disagree) / 1 (neutral) /2 (agree) /3 (Strongly agree)	
Focus group	Changed the initial item structure to three declarative statements. "best practice" was changed to "evidence-based practice"			
Version 2	My [organization/clinical setting] supports evidence-based practice.	My [organization/clinical setting] somewhat supports evidence-based practice.	My [organization/clinical setting] does not supports evidence-based practice.	
Research team	Changed from "my org supports best practice" to "I feel that I have the necessary resources to …"			
Version 3	I feel that I have the necessary resources to integrate research evidence into my practice.	I feel that I only have some of the necessary resources to integrate research evidence into my practice.	I feel that I do not have the necessary resources to integrate research evidence into my practice.	
Cognitive interviews	Omitted "I feel that". Omitted "I only have some". Changed "I do not have" for "I have few". Added the examples in parentheses for resources			
Version 4	I **have the necessary resources** (e.g. (paid time to consult the evidence, access to journals,	I **have some of the necessary resources** to integrate research evidence into my practice.	I **have few of the necessary resources** to integrate research evidence into my practice.	

	therapeutic material...) to integrate research evidence into my practice.			
Cognitive interviews	Added an asterisk with examples of resources			
Version 5	I have **the necessary resources*** to integrate research evidence into my practice.	I have **some of the necessary resources*** to integrate research evidence into my practice.	I have **few of the necessary resources*** to integrate research evidence into my practice.	*Examples: paid time to consult the evidence, access to journals, therapeutic material...
Version 6	Idem			
Cognitive interviews	Added computer to the example			
Version 7	Idem			*Examples: paid time to consult the evidence, access to journals, access to computer, therapeutic material...
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			
Version 11	Idem			
Version 12	Idem			
Item: Attitudes → Effort				
Version 1	Please indicate your level of agreement with the following statements:	I am willing to use new and different types of clinical interventions (e.g., assessment, treatment) developed by researchers to help my patients/clients.	0 (Strongly disagree)/1 (Disagree)/ 2 (Neutral)+ 3 (Agree) / 4 (Strongly agree)	
Focus group	Changed the initial item structure to three declarative statements. Replaced the ending "to help my patients/clients" with "for my patient's care plan". Replaced "to use new and different types of clinical interventions (e.g. assessment, treatment) developed by researchers" with "to use EBP" to simplify the item.			

Version 2	I am [willing/inclined] to use EBP for my patient's* care plan	I am [somewhat willing/somewhat inclined] to use EBP for my patient's* care plan	I am [not willing/not inclined] to use EBP for my patient's* care plan	
Research team	Changed from "I am willing to use EBP…" to the notion of "worth the effort …" to diminish social desirability bias.			
Version 3	Incorporating evidence into my practice is definitely worth the effort.	Incorporating evidence into my practice is somewhat worth the effort.	Incorporating evidence into my practice is not worth the effort.	
Cognitive interviews	Switched the structure of the sentence. Added "(not) really (worth the effort)". Added parentheses with specification after practice.			
Version 4	It is **definitely worth the effort** to incorporate research evidence into my practice (i.e. assessment and/or intervention plan).	It is **somewhat worth the effort** to incorporate research evidence into my practice.	It is **not really worth the effort** to incorporate research evidence into my practice.	
Cognitive interviews	Changed "incorporate" for "integrate" for consistency throughout the measure. Deleted the example because "practice is used before without specification". Changed "worth THE effort" for "worth MY effort".			
Version 5	It is **definitely worth my effort** to integrate research evidence into my practice.	It is **somewhat worth my effort** to incorporate research evidence into my practice.	It is **not really worth my effort** to incorporate research evidence into my practice.	
Version 6	Idem			
Cognitive interviews	Removed the superlative "definitely" from first response option. Changed "my (effort)" for "the (effort)"			
Version 7	It is **worth** the effort to integrate research evidence into my practice.	It is **somewhat** worth the effort to incorporate research evidence into my practice.	It is **not really** worth the effort to incorporate research evidence into my practice.	
Cognitive interviews	Changed the focus on the idea of 'worth' for the concept of 'requiring effort' to decrease social desirability bias.			
Version 8	It requires **little effort** to integrate research evidence into practice	It requires **some effort** to integrate research evidence into practice	It requires **a lot of effort** to integrate research evidence into practice	
Cognitive interviews	changed "some" effort to "moderate" effort			

Version 9	It requires **little effort** to integrate research evidence into practice.	It requires **moderate effort** to integrate research evidence into practice.	It requires **a lot of effort** to integrate research evidence into practice.	
Version 10	Idem			
Cognitive interviews	Added “(effort) for me”			
Version 11	It requires **little effort** for me to integrate research evidence into practice.	It requires **moderate effort** for me to integrate research evidence into practice.	It requires **a lot of effort** for me to integrate research evidence into practice.	
Version 12	Idem			
Item: Activities related to EBP →Keep up to date				
Version 1	In the past month, how often have you: made time to read research?	0 (Never) / 1 (Monthly or less) / 2 (Bi-weekly) / 3 (Weekly). /4 (Daily)		
Focus group	Changed to declarative statements. Omitted the idea of making time to read research. Changed “reading research” to “consulting research evidence” to be more inclusive to other sources of research evidence			
Version 2	I regularly consult research evidence.	I occasionally consult research evidence.	I rarely consult research evidence.	
Version 3	Idem			
Version 4	Idem			
Version 5	Idem			
Version 6	Idem			
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Cognitive interviews	Changed "consult research evidence" for "keep up to date with research evidence" because participants were interpreting the item as “frequency of using evidence in their practice” and the activity of keeping up with knowledge of scientific research outside of their clinical encounters			
Version 10	I **regularly** keep up to date with research evidence.	I **occasionally** keep up to date with research evidence.	I **rarely** keep up to date with research evidence.	
Version 11	Idem			
Version 12	Idem			
Iterations of the instructions and changes to the visual presentation (in italics)				
Version 1	None			

Version 2	None			
Research team	Developed and added instructions for the index			
Version 3	For each group of statements, select ONE statement which best applies to you. Please respond as honestly as possible.			
Cognitive interviews	Removed the last sentence from the previous version. Specified "best reflects your current practice" to contextualize the answers. Added the word "instructions". *Bolded the response options for each item.*			
Version 4	Instructions: For each group of statements, please select ONE statement which best reflects your current practice.			
Cognitive interviews	Added "(best reflects your current practice) and context"			
Version 5	Instructions: For each group of statements, please select ONE statement which best reflects your current practice and context.			
Cognitive interviews	Replaced "For each group of statements" with "from each box" and *added boxes around each of the five items. Supplemented the checkboxes for each response choice with numbers and letters.*			
Version 6	Please **select ONE** statement **from each box** which best reflects your current practice and context.			
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			
Version 11	Idem			
Version 12	Idem			

French version of the index

Question: Utilisation autodéclarée des données probantes				
Version 1	Depuis un mois, à quelle fréquence avez-vous...	Décider d'un plan d'action approprié intégrant des données probantes, le jugement clinique et les préférences du client ou patient?	0 (Jamais) / 1 (Une fois ou plus)	
Groupe de discussion	Changed the initial item structure to three declarative statements. Changed "[integrating the 3 pillars]" to « *intégrer les évidences scientifiques* ». Changed « *décider d'un plan d'action approprié* » to « *…dans le plan de soins* ».			
Version 2	J'intègre régulièrement les évidences scientifiques dans le plan de soins.	J'intègre occasionnellement les évidences scientifiques dans le plan de soins.	J'intègre rarement ou jamais les évidences scientifiques dans le plan de soins.	

Équipe de recherche	Changed item to elicit the first source of information when faced with a practice uncertainty.			
Version 3	Face à une incertitude dans ma pratique, je m'appuie sur les évidences scientifiques.	Face à une incertitude dans ma pratique, je m'appuie sur mes collègues.	Face à une incertitude dans ma pratique, je m'appuie sur mon expérience clinique.	
Version 4	Idem			
Version 5	Idem			
Entretiens cognitifs	Added an asterisk to define what we mean by « *incertitude dans ma pratique* ».			
Version 6	Je m'appuie premièrement sur **les évidences scientifiques** lorsque je fais face à une incertitude dans ma pratique.	Je m'appuie premièrement sur **mes collègues** lorsque je fais face à une incertitude dans ma pratique.	Je m'appuie premièrement sur **mon expérience clinique** lorsque je fais face à une incertitude dans ma pratique.	**incertitude: une situation dans laquelle il existe un déficit dans vos connaissances en lien avec une décision clinique*
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			
Entretiens cognitifs	To decrease variability due to clinical population and context, the item has been modified to focus on frequency of using research evidence when faced with a practice uncertainty.			
Version 11	Lorsque je fais face à une incertitude* dans ma pratique, j'utilise **presque toujours** les données probantes.	j'utilise **parfois** les données probantes.	j'utilise **rarement** les données probantes.	**incertitude: une situation dans laquelle il existe une lacune dans vos connaissances concernant une décision clinique.*
Version 12	Idem			
Question: Auto-efficacité				
Version 1	Veuillez indiquer à quel point vous avez confiance en vos capacités actuelles en choisissant le nombre correspondant sur l'échelle d'appréciation suivante.	Déterminer si des preuves découlant d'une recherche de la littérature s'appliquent à la situation de votre patient ou client?	0 (Aucune confiance à 25%)) / 1 (25-50%) / 2 (50-75%) /3 (75% à confiance totale)	
Groupe de discussion	Changed the initial item structure to three declarative statements. Omitted « *confiance en vos capacités actuelles* » to shorten the sentence. Changed « *preuves découlant d'une recherche de la littérature* » to be inclusive to other souces of research evidence.			

Version 2	Je suis confiant.e de décider si les évidences scientifiques s’appliquent à mes patients*	Je suis plutôt confiant.e de décider si les évidences scientifiques s’appliquent à mes patients*.	Je ne suis pas confiant.e de décider si les évidences scientifiques s’appliquent à mes patients*.	*Patient et client sont utilisés comme synonymes
Équipe de recherche	Changed « décider si les évidences scientifiques s’appliquent à mes patients » to « *d'intégrer les évidences scientifiques dans mon plan d'intervention* ». Added « très » superlative to first and « peu » last level.			
Version 3	Je suis très confiant(e) en mes capacités d'intégrer les évidences scientifiques dans mon plan d'intervention.	Je suis moyennement confiant(e) en mes capacités d'intégrer les évidences scientifiques dans mon plan d'intervention.	Je suis peu confiant(e) en mes capacités d'intégrer les évidences scientifiques dans mon plan d'intervention.	
Version 4	Idem			
Entretiens cognitifs	Changed « *d'intégrer* » for « *d'appliquer* ». Changed « *plan d'intervention* » for « *cas cliniques* ».			
Version 5	Je suis **très confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	Je suis **moyennement confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	Je suis **peu confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	
Version 6	Idem			
Version 7	Idem			
Version 8	Idem			
Entretiens cognitifs	Changed « *très confiant(e)* » to « *confiant(e)* »			
Version 9	Je suis **confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	Je suis **moyennement confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	Je suis **peu confiant(e)** en mes capacités d'appliquer les évidences scientifiques à des cas cliniques.	
Entretiens cognitifs	Changed « *cas cliniques* » for « *en pratique* ».			
Version 10	Je suis **confiant(e)** en mes capacités d'appliquer les données probantes en pratique.	Je suis **moyennement confiant(e)** en mes capacités d'appliquer les données probantes en pratique.	Je suis **peu confiant(e)** en mes capacités d'appliquer les données probantes en pratique.	
Version 11	Idem			
Version 12	Idem			
Question: Ressources				

Version 1	Veuillez indiquer à quel point vous êtes en accord avec les énoncés suivants.	Mon organisation soutient les pratiques optimales.	0 (Tout à fait en désaccord /En désaccord)) / 1 (Neutre) /2 (D'accord) /3 (Tout à fait d'accord)	
Groupe de discussion	Changed the initial item structure to three declarative statements. « *pratiques optimales* » was changed to « *la pratique basée sur les données probantes* »			
Version 2	Mon [organisation/milieu de clinique] soutient la pratique basée sur les données probantes.	Mon [organisation/milieu de clinique] soutient plus ou moins la pratique basée sur les données probantes.	Mon [organisation/milieu de clinique] ne soutient pas la pratique basée sur les données probantes.	
Équipe de recherche	Changed from « *Mon [organisation/milieu de clinique] soutient …*» to « *J'estime avoir les ressources nécessaires …*»			
Version 3	J'estime avoir les ressources nécessaires pour intégrer les évidences scientifiques dans ma pratique.	J'estime avoir seulement une partie des ressources nécessaires pour intégrer les évidences scientifiques dans ma pratique.	J'estime ne pas avoir les ressources nécessaires pour intégrer les évidences scientifiques dans ma pratique.	
Version 4	Idem			
Entretiens cognitifs	Omitted « *J'estime (avoir)* ». Changed « *ne pas avoir les ressources…*» for « **(***avoir) peu de ressources nécessaires*». Added an asterisk with examples of resources.			
Version 5	J'ai **les ressources nécessaires** pour intégrer les évidences scientifiques dans ma pratique.	J'ai **seulement une partie des ressources nécessaires** pour intégrer les évidences scientifiques dans ma pratique.	J'ai **peu de ressources nécessaires** pour intégrer les évidences scientifiques dans ma pratique.	**Exemples : du temps payé pour lire les données probantes, accès aux journaux, accès au matériel thérapeutique nécessaire*
Version 6	Je possède **les ressources* nécessaires** pour intégrer les données probantes dans ma pratique.	Je possède **une partie des ressources* nécessaires** pour intégrer les données probantes dans ma pratique.	Je possède **peu de ressources* nécessaires** pour intégrer les données probantes dans ma pratique.	**Exemples : du temps payé pour lire les données probantes, accès aux journaux, accès au matériel thérapeutique nécessaire*

Entretiens cognitifs	Added computer to the example			
Version 7	Idem			**Exemples : du temps payé pour lire les données probantes, accès aux journaux, accès à un ordinateur, accès au matériel thérapeutique nécessaire*
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			
Version 11	Idem			
Version 12	Idem			
Question: Attitudes → Effort				
Version 1	Veuillez indiquer à quel point vous êtes en accord avec les énoncés suivants.	J'accepterais de bon gré d'utiliser divers types d'interventions cliniques inédites (ex. évaluation, traitement) mises au point par des chercheurs pour aider mes patients ou clients.	0 (Tout à fait en désaccord /1 (En désaccord)) / 2 (Neutre) /3 (D'accord) /4 (Tout à fait d'accord)	
Groupe de discussion	Changed the initial item structure to three declarative statements. Replaced the ending « *pour aider mes patients ou clients* » with « *pour le plan de soins de mes patients* ». Replaced « *d'utiliser divers types d'interventions cliniques inédites (ex. évaluation, traitement) mises au point par des chercheurs* » with « *utiliser des données probantes* » to simplify the item.			
Version 2	Je suis [prêt à/partant de/tenté de] utiliser des données probantes pour le plan de soins de mes patients*.	Je suis plus ou moins [prêt à/partant de/tenté de] utiliser des données probantes pour le plan de soins de mes patients*.	Je ne suis pas [prêt à/partant de/tenté de] utiliser des données probantes pour le plan de soins de mes patients*.	
Équipe de recherche	Changed from « *Je suis prêt à utiliser des données probantes…*» to the notion of « *en valloir la peine* » to diminish social desirability bias.			
Version 3	Intégrer les évidences scientifiques dans ma pratique en vaut vraiment la peine.	Intégrer les évidences scientifiques dans ma pratique en vaut plus ou moins la peine.	Intégrer les évidences scientifiques dans ma pratique n'en vaut pas la peine.	
Version 4	Idem			

Entretiens cognitifs	Inversed the structure of the sentence. Added « *n'en vaut pas <u>vraiment</u> la peine* ». Added «… *la peine <u>pour moi</u>* ».			
Version 5	Cela **vaut vraiment la peine** pour moi d'intégrer les évidences scientifiques dans ma pratique.	Cela **vaut plus ou moins la peine** pour moi d'intégrer les évidences scientifiques dans ma pratique.	Cela **n'en vaut pas vraiment la peine** pour moi d'intégrer les évidences scientifiques dans ma pratique.	
Version 6	Idem			
Entretiens cognitifs	Changed « (valloir) *la peine <u>pour moi</u>* » for « (valloir) *l'effort* » Changed the focus on the idea of 'worth' for the concept of 'requiring effort' to decrease social desirability bias. Removed the superlative « *vraiment* » from first response option.			
Version 7	Cela **vaut l'effort** d'intégrer les évidences scientifiques dans ma pratique.	Cela **vaut moyennement l'effort** d'intégrer les évidences scientifiques dans ma pratique.	Cela **ne vaut pas vraiment l'effort** d'intégrer les évidences scientifiques dans ma pratique.	
Entretiens cognitifs	Replaced « (valloir) l'effort » for « nécessite de l'effort »			
Version 8	Cela **nécessite peu d'efforts** pour intégrer les évidences scientifiques dans ma pratique.	Cela **nécessite quelques efforts** pour intégrer les évidences scientifiques dans ma pratique.	Cela **nécessite beaucoup d'efforts** pour intégrer les évidences scientifiques dans ma pratique.	
Entretiens cognitifs	Changed « *quelques efforts* » to « *effort modéré* »			
Version 9	Cela **nécessite peu d'effort** pour intégrer les évidences scientifiques dans ma pratique.	Cela **nécessite un effort modéré** pour intégrer les évidences scientifiques dans ma pratique.	Cela **nécessite beaucoup d'effort** pour intégrer les évidences scientifiques dans ma pratique.	
Version 10	Idem			
Entretiens cognitifs	Starting the sentence with "cela" was problematic in French; sentence structure was inversed from English. Added "me (demande)" to specify that the focus is on the respondent's perceived effort.			
Version 11	Intégrer les données probantes dans ma pratique me demande **peu d'effort.**	Intégrer les données probantes dans ma pratique me demande **un effort modéré**.	Intégrer les données probantes dans ma pratique me demande **beaucoup d'effort**.	
Version 12	Idem			
Question: Activités reliées à la PFDP → Se tenir à jour				
Version 1	Depuis un mois, à quelle fréquence avez-vous...	Réservé du temps à la lecture de travaux de recherche?	0 (Jamais) / 1 (Une fois par mois ou moins) / 2 (Aux 2 semaines) / 3	

			(Toutes les semaines) / 4 (Tous les jours)	
Groupe de discussion	Changed to declarative statements. Omitted the idea of making time to read research. Changed « *la lecture de travaux de recherche* » to « *consulter les évidences scientifiques* » to be more inclusive to other sources of research evidence.			
Version 2	Je consulte régulièrement les évidences scientifiques.	Je consulte occasionnellement les évidences scientifiques.	Je consulte rarement ou jamais les évidences scientifiques.	
Version 3	Idem			
Version 4	Idem			
Version 5	Idem			
Version 6	Idem			
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Entretiens cognitifs	Changed to « *consulter les évidences scientifiques* » for « *(se tenir) à jour quant aux données probantes* » because participants were interpreting the item as "frequency of using evidence in their practice" and the activity of keeping up with knowledge of scientific research outside of their clinical encounters			
Version 10	Je me tiens **régulièrement** à jour quant aux données probantes.	Je me tiens **occasionnellement** à jour quant aux données probantes.	Je me tiens **rarement** à jour quant aux données probantes.	
Version 11	Idem			
Version 12	Idem			
Iterations of the instructions and changes to the visual presentation (in italics)				
Version 1	None			
Version 2	None			
Équipe de recherche	Developed and added instructions for the index			
Version 3	Pour chaque groupe d'énoncés, veuillez sélectionner UN énoncé qui s'applique le mieux à vous. Merci de répondre le plus honnêtement possible.			
Entretiens cognitifs	Removed the last sentence from the previous version. Specified « *(s'applique le mieux à) votre pratique* » to contextualize the answers. Added the word "*instructions*". *Bolded the response options for each item.*			
Version 4	Instructions: Pour chaque groupe d'énoncés, veuillez sélectionner UN énoncé qui reflète le mieux votre pratique.			
Entretiens cognitifs	Added « *(s'applique le mieux à) votre pratique et contexte actuel* »			
Version 5	Instructions: Pour chaque groupe d'énoncés, veuillez sélectionner UN énoncé qui reflète le mieux votre pratique et contexte actuel.			

Entretiens cognitifs	Replaced « *Pour chaque groupe d'énoncés* » with « *de chaque case* ». *Added boxes around each of the five items. Supplemented the checkboxes for each response choice with numbers and letters. Bolded key words in the instructions.*			
Version 6	Veuillez **sélectionner UN** énoncé **de chaque case** qui reflète le mieux votre pratique et contexte actuel.			
Version 7	Idem			
Version 8	Idem			
Version 9	Idem			
Version 10	Idem			
Version 11	Idem			
Version 12	Idem			

Appendix III. Overview of the item evolution process during cognitive interviews

Participant #	Language	Profession	Index version	Use of RE	Self-efficacy	Resources	Attitudes	Activities related to EBP	Instructions	Visual
P1	EN	PT	3	Major	Minor	Major	Minor		Minor	Minor
P2	EN	PT	3	Minor			Minor		Minor	
P3	EN	OT	3	Major		Minor	Minor		Minor	
P4	FR	OT	3	Minor	Major	Minor	Minor		Minor	
New version 4				**M**		**M**	**M**		**M**	**M**
P5	EN	OT	4		Minor	Minor				
New version 5					**M**	**M**	**M**		**M**	
P6	FR	OT	5	Minor		Minor				
P7	EN	OT	5	Minor					Minor	Minor
P8	EN	PT	5	Minor	Minor		Major		Minor	
New version 6				**M**					**M**	**M**
P9	FR	PT	6				Minor	Minor		
P10	FR	OT	6			Minor	Minor			
New version 7						**M**	**M**			
P11	FR	PT	7		Minor		Major			
P12	FR/EN	OT	7		Minor	Minor	Minor			
P13	FR	PT	7	Minor			Major			
New version 8							**M**			
P14	EN	PT	8	Minor			Minor			
P15	EN	OT	8		Minor		Minor			
New version 9					**M**		**M**			
P16	EN	OT	9		Minor			Minor		
P17	FR	PT	9				Minor	Minor		
New version 10					**M**			**M**		
P18	FR	OT	10							
P19	FR	OT	10	Major			Minor	Minor		
P20	FR	OT	10	Minor	Minor	Minor				
P21	EN	OT	10	Minor						
New version 11				**M**			**M**			
P22	FR	PT	11							
P23	EN	OT	11							
P24	EN	PT	11					Minor		
Version 12 accepted										
Total iterations during cognitive interviews				3	3	3	6	1		

EN: English; FR: French; PT: physical therapist; OT: occupational therapist; RE: research evidence; Major: Major issue identified by participants; Minor: minor issue identified by participants; M: Modification implemented by the research team into next version of the index

Appendix IV. Final index in English and French

Propensity to Integrate Research Evidence into Clinical Decision-Making Index (PIRE-CDMI)

Please **select ONE** statement **from each box** which best reflects your current practice and context.

1. When faced with a practice uncertainty*,

☐ a) I **almost always** use research evidence.

☐ b) I **sometimes** use research evidence.

☐ c) I **rarely** use research evidence.

**uncertainty: a situation in which there is a gap in your knowledge relating to a clinical decision*

2.

☐ a) I am **confident** in my ability to apply research evidence to practice.

☐ b) I am **somewhat confident** in my ability to apply research evidence to practice.

☐ c) I am **not very confident** in my ability to apply research evidence to practice.

3.

☐ a) I have **the necessary resources*** to integrate research evidence into my practice.

☐ b) I have **some of the necessary resources*** to integrate research evidence into my practice.

☐ c) I have **few of the necessary resources*** to integrate research evidence into my practice.

**Examples: paid time to consult the evidence, access to journals, access to a computer, access to necessary therapeutic material...*

4.

☐ a) It requires **little effort** for me to integrate research evidence into practice.

☐ b) It requires **moderate effort** for me to integrate research evidence into practice.

☐ c) It requires **a lot of effort** for me to integrate research evidence into practice.

5.

☐ a) I **regularly** keep up to date with research evidence.

☐ b) I **occasionally** keep up to date with research evidence.

☐ c) I **rarely** keep up to date with research evidence.

Indice de la tendance à intégrer les données probantes dans la prise de décisions cliniques (I-TIDP-PDC)

Veuillez **sélectionner UN** énoncé **de chaque encadré** qui reflète le mieux votre pratique et contexte actuel.

1. Lorsque je fais face à une incertitude* dans ma pratique,

☐	a) J'utilise **presque toujours** les données probantes.
☐	b) J'utilise **parfois** les données probantes.
☐	c) J'utilise **rarement** les données probantes.
	**incertitude: une situation dans laquelle il existe une lacune dans vos connaissances concernant une décision clinique.*

2.

☐	a) Je suis **confiant(e)** en mes capacités d'appliquer les données probantes en pratique.
☐	b) Je suis **moyennement confiant(e)** en mes capacités d'appliquer les données probantes en pratique.
☐	c) Je suis **peu confiant(e)** en mes capacités d'appliquer les données probantes en pratique.

3.

☐	a) Je possède **les ressources* nécessaires** pour intégrer les données probantes dans ma pratique.
☐	b) Je possède **une partie des ressources* nécessaires** pour intégrer les données probantes dans ma pratique.
☐	c) Je possède **peu de ressources* nécessaires** pour intégrer les données probantes dans ma pratique.
	**Exemples : du temps payé pour lire la littérature, accès aux journaux, accès à un ordinateur, accès au matériel thérapeutique nécessaire*

4.

☐	a) Intégrer les données probantes dans ma pratique me demande **peu d'effort**.
☐	b) Intégrer les données probantes dans ma pratique me demande un **effort modéré**.
☐	c) Intégrer les données probantes dans ma pratique me demande **beaucoup d'effort**.

5.

☐	a) Je me tiens **régulièrement** à jour quant aux données probantes.
☐	b) Je me tiens **occasionnellement** à jour quant aux données probantes.
☐	c) Je me tiens **rarement** à jour quant aux données probantes.

Appendix V. Regression estimates of the effect of language and response option on score per item

MODEL: Score (0-100) = language (English, French) + response option (Low, Middle, High) + interaction terms								
Self-efficacy								
		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B	
		Beta	Std. Error	Beta	t		Lower Bound	Upper Bound
1	Referent (French, Not very confident)	13.6	0.9		14.8	<.001	11.8	15.4
	English	0.5	1.3	0.0	0.4	0.688	-2.1	3.2
	Somewhat confident	37.7	1.3	0.6	29.1	<.001	35.1	40.2
	Confident	72.7	1.3	1.1	56.1	<.001	70.1	75.2
	English*Somewhat confident	-2.0	1.9	0.0	-1.1	0.287	-5.7	1.7
	English*Confident	-1.4	1.9	0.0	-0.7	0.457	-5.1	2.3
	Residual statistics	MIN	MAX					
	Std. Residual	-3.334	3.831					
	Cook's Distance	0	0.061					
Use of research evidence								
		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B	
		Beta	Std. Error	Beta	t		Lower Bound	Upper Bound
1	Referent (French, Rarely)	11.9	1.0		12.4	<.001	10.0	13.8
	English	0.0	1.4	0.0	0.0	0.992	-2.7	2.8
	Sometimes	35.5	1.4	0.5	26.1	<.001	32.8	38.2
	Almost always	75.3	1.4	1.1	55.4	<.001	72.6	78.0
	English*Sometimes	0.7	2.0	0.0	0.4	0.719	-3.2	4.6

	English*Almost always	0.3	2.0	0.0	0.2	0.871	-3.6	4.2
	Residual statistics	MIN	MAX					
	Std. Residual	-3.592	4.316					
	Cook's Distance	0	0.086					
Effort								
		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B	
		Beta	Std. Error	Beta	t		Lower Bound	Upper Bound
1	Referent (French, Little effort)	19.0	0.8		24.0	<.001	17.4	20.6
	English	-0.4	1.2	0.0	-0.3	0.749	-2.6	1.9
	Moderate effort	34.3	1.1	0.6	30.6	<.001	32.1	36.5
	A lot of effort	66.6	1.1	1.1	59.4	<.001	64.4	68.8
	English*Moderate effort	-0.6	1.6	0.0	-0.4	0.702	-3.8	2.6
	English*A lot of effort	2.0	1.6	0.0	1.2	0.224	-1.2	5.2
	Residual statistics	MIN	MAX					
	Std. Residual	-3.696	3.242					
	Cook's Distance	0	0.057					
Keep up to date								
		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B	
		Beta	Std. Error	Beta	t		Lower Bound	Upper Bound
1	Referent (French, Rarely)	4.9	2.0		2.5	0.013	1.1	8.8
	English	-0.3	2.9	0.0	-0.1	0.924	-5.9	5.3
	Occasionally	16.8	2.8	0.4	6.1	<.001	11.4	22.3

	Regularly	41.9	2.8	0.9	15.1	<.001	36.4	47.4
	English*Occasionally	-0.8	4.0	0.0	-0.2	0.848	-8.7	7.2
	English*Regularly	-1.7	4.0	0.0	-0.4	0.668	-9.7	6.2
	Residual statistics	MIN	MAX					
	Std. Residual	-2.631	4.335					
	Cook's Distance	0	0.087					
Resources								
		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B	
		Beta	Std. Error	Beta	t		Lower Bound	Upper Bound
1	Referent (French, Few of the necessary resources)	14.24	0.94		15.18	<.001	12.4	16.1
	English	-0.21	1.36	0.00	-0.16	0.876	-2.9	2.5
	Some of the necessary resources	28.67	1.33	0.50	21.61	<.001	26.1	31.3
	The necessary resources	64.60	1.33	1.12	48.70	<.001	62.0	67.2
	English*Some of the necessary resources	1.15	1.92	0.02	0.60	0.551	-2.6	4.9
	English*The necessary resources	0.91	1.92	0.01	0.47	0.639	-2.9	4.7
	Residual statistics	MIN	MAX					
	Std. Residual	-3.923	3.482					
	Cook's Distance	0	0.071					

Chapter 8: The Integration of Manuscripts 4 and 5

Research objectives of manuscripts 4 and 5

Manuscript 4:

The global aim of this study was to contribute evidence for the clarity and interpretability of items and response options for a new bilingual (English and French) measure, *the Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). Specifically, the primary objective of this research was to qualitatively review and revise the included items in the prototype index in English and French The secondary objective was to estimate the equivalency of response option labels in both languages.

Manuscript 5:

The global aim of this study was to develop a scoring algorithm for the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific objective was to estimate the part-worth utilities and relative importance of five key dimensions to the propensity to integrate research evidence into clinical decision-making, information that will be used in the scoring algorithm. The secondary objective was to contribute evidence of the ability of the PIRE-CDMI to distinguish between known groups of OTs and PTs and to compare with a global self-rating of EBP.

Integration of manuscripts 4 and 5

Manuscript 4 described the three-phased rewriting process of the items included in the prototype index. A focus group (phase 1) and 24 cognitive interviews (phase 2) allowed for an examination of the cognitive process involved when end-users respond to items (i.e., evidence based on response process[1]). This process led to 12 iterations of changes to the index to optimise the clarity and interpretability of items and ultimately, reduce measurement error in the index and cognitive burden for respondents. Among numerous minor changes, three major modifications were implemented to ensure that the items were best representative of the actual process of integrating research evidence into CDM that occurs in clinical practice. Throughout the item rewriting process, modifications to items suggested in one language led to the equivalent modification in the other language. The revised items were then tested in subsequent cognitive interviews. In phase 3 of this study, response options were found to be equivalent in both languages through a cross-sectional survey of native English and French speakers.

The final study in this dissertation corresponds to the last phase of measure development (*d*), development of procedures and materials for administration and scoring[1]. The research presented in *Manuscript 5* describes the procedures required to develop a weighted scoring algorithm for the index. Such a scoring system is required to generate a total score that is representative of the relative importance of the five included domains, as these domains may contribute to different degrees to a clinician's propensity to integrate research evidence into CDM. In this study, part-worth utilities for each dimension-level are estimated using a best-worst scaling choice exercise using state-of-the art software, *Sawtooth*[2]. This is therefore the first attempt at identifying the quantitative importance of each of the five EBP domains on a clinician's propensity to integrate research evidence into CDM.

Manuscript 5 also contributes to the evaluation of items which is associated with the second phase of measure development (*b*)[1]. This study provides evidence regarding the interpretability of scores, specifically relationships between index scores and profession (OT and PT) and with a global self-rating score of EBP.

References

1. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing*.; 2014.

2. Howell J. CBC/HB for Beginners. Sawtooth Software. Published 2009. Accessed April 20, 2022. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-for-beginners

Chapter 9: Manuscript 5

Propensity to Integrate Research Evidence into the Decision-making Process in Rehabilitation: Development of a Weighted Algorithm Using a Best-Worst Scaling Choice Exercise

Jacqueline Roberge-Dao[1,2], Aliki Thomas[1,2,3], Annie Rochette[2,4], Keiko Shikako[1,2], Nancy Mayo[1,5]

[1] School of Physical and Occupational Therapy, McGill University, Montréal, Canada
[2] Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Canada
[3] The Institute of Health Sciences Education, McGill University, Montréal, Canada
[4] L'École de réadaptation de l'Université de Montréal, Montréal, Canada
[5] Center for Outcomes Research and Evaluation (CORE), The Research Institute of the McGill University Health Center (RI-MUHC), Montreal, QC, Canada

Abstract

Objectives: The factors underpinning a rehabilitation clinician's propensity to integrate research evidence into the decision-making process are multifaceted. While studies have focused on these individual and organizational factors individually, little is known about how they combine and interact. The global aim of this study was to develop a scoring algorithm for the Propensity to Integrate Research Evidence into Clinical Decision-Making Index (PIRE-CDMI), a brief multidimensional index.

Methods: A cross-sectional online survey using Sawtooth software was conducted to gather choice-based data from occupational and physical therapists, clinical managers, and experts in EBP. Using the best-worst scaling (multi-profile case), a type of choice experiment, respondents were asked to select "the best" and "the worst" among three hypothetical profiles containing varied levels of the five EBP dimensions (fractional factorial and balanced overlap design). Respondents completed 12 rounds of this task (randomized draws) and answered questions related to acceptability. Part-worth utilities were generated for all dimension-levels using Hierarchical Bayes analysis.

Results: A total of 61 responses were used for analysis after eliminating individuals who did not complete the survey (n=20) and misunderstood the task (n=5). Part-worth utilities were integrated into a scoring algorithm allowing for the generation of a total score for each index respondent. Based on average relative importance, participants highly valued two domains - *use of research evidence* (35% average relative importance) and *keep up to date* (26% average relative importance), relative to the other three domains - *self-efficacy* (14% average relative importance), *effort* (13% average relative importance) and *resources* (12% average relative importance).

Conclusion: The results of this study provide a strong basis for estimating the numerical influence of five EBP dimensions on a rehabilitation clinician's propensity to integrate research evidence into their decision-making. The newly developed scoring algorithm based on the part-

worth utilities allows for the development of a brief multidimensional index in rehabilitation with a mathematically and theoretically valid total score.

Introduction

Given the necessity for effective and efficient healthcare, evidence-based practice (EBP) has become the basis for clinical practice across the healthcare professions including occupational and physical therapy. In their pursuit to optimize function and quality of life, occupational (OTs) and physical therapists (PTs) are expected to integrate research evidence into their clinical decision-making process.

EBP is well-recognized as the integration of best research evidence with clinical expertise and patient preferences and values within the context of available resources[1,2]. In an effort to enhance quality of care and health outcomes, professional bodies have advocated for EBP through their position statements[3,4]. In turn, professional entry-level programs have integrated EBP content into their curricula. Still, despite aggregated efforts to equip OTs and PTs with the competencies associated with becoming evidence-based clinicians, studies demonstrate that the uptake of research evidence is lagging and inconsistent[5–10].

In the last two decades, there has been a surge of interest in studying the factors that influence a rehabilitation clinician's ability to successfully exercise EBP. Several studies have demonstrated that a set of individual and organizational factors influence the enactment of EBP. These include lack of protected time for activities related to EBP (e.g., reading articles), low self-efficacy in participating in EBP-related activities, lack of skills (e.g., inability to understand statistics, critically appraise scientific papers), insufficient knowledge on evidence-based interventions, limited resources in the practice setting (e.g., lack of access to journals; inadequate equipment to implement evidence-based interventions), and difficulty forming new habits using novel interventions[5,8,11–13]. Prior research has provided valuable insights into (1) the factors to prioritize when supporting clinicians' engagement with the EBP process and (2) the necessity of better tailoring EBP interventions to maximize relevance and EBP outcomes.

What we know about the factors[j] that influence EBP and the effectiveness of EBP interventions is largely derived from self-report questionnaires measuring one or a combination of EBP domains[k]. In rehabilitation, systematic reviews have identified 34 (OTs) and 24 (PTs)

[j] In this study, *factor* is used to describe an individual or organizational influence on a clinician's propensity to integrate research evidence into clinical decision-making.

[k] *Domain* is used in the context of measurement to describe a conceptually defined part of a construct. *Domain* can also be used to represent a subscale of a measure[14].

questionnaires[15,16]; since then, two other measures have been developed for use with rehabilitation clinicians[17,18]. Among the questionnaires covering multiple EBP domains in rehabilitation, these domains are described separately[6,17,19–30]. These multidimensional EBP questionnaires can be considered as *profiles* whereby results are summarized and interpreted by sub-scale, where each sub-scale represents an EBP domain[14]. With profiles, the relative importance of one domain in relation to others is unaccounted for resulting in a fragmented interpretation of EBP.

As an example, the *Evidence-based Practice Profile questionnaire* (EBP^2) by McEvoy et al.[31] includes five EBP domains (Relevance, Terminology, Confidence, Practice and Sympathy) and a domain score is calculated by adding all of the items in each domain. However, it is unclear how the different domains interact with each other to impact EBP[32]. For instance, how does a clinician's knowledge of EBP concepts (Terminology domain) interact with their self-reported confidence (Confidence domain) in influencing their likelihood of engaging in the EBP process? Despite gaining in-depth information on single domains, the information derived from these measures is limited because of the omnipresence of multiple factors in practice. There is a need to identify the degree to which each factor impacts clinicians relative to the other factors to provide a more nuanced picture of EBP.

One avenue that may be pursued to determine the relationships between EBP factors is quantification of the relative influence of each factor by estimating weights. Various methods of generating weights or utilities have been used in the health-related quality of life[33], patient preference[34] and health economics[35] literature. In the field of health-related quality of life, preference-based measures have become widely recognized for providing a single total score representative of the overall health state whilst considering the relative weight or value of each dimension[l] [37–39]. Preference-based measures are brief, typically consisting of one item per domain, and can be generic or condition specific (e.g., multiple sclerosis, stroke, and obesity)[40–42].

To determine the relative weights of each dimension contained within a preference-based measure, individuals are asked to provide input on the relative importance of certain domains compared to others (i.e., their preferences) using one or more weight elicitation techniques[33–35].

[l] In this context of preference-based measures or multidimensional indices, *dimension* describes the selected item best representing the domain [36].

Broadly, two main approaches are used to determine weights. The first approach involves ranking, rating, or choice designs (e.g., conjoint analysis). This approach, which was developed in marketing research to predict buyer behavior, aims to explore how changes in product attributes[m] can impact choice[43,44]. The second approach consists of methods using direct elicitation of numerical or monetary values (e.g., willingness-to-pay or rating scale). This approach aims to estimate the demand for a specific attribute within a product [44]. The selection of a specific method depends on the fit with the study perspective.

Although these weight elicitation methods have not previously been used in EBP studies, scholars in the related field of implementation science[45] have specifically endorsed conjoint analysis as an innovative method to determine the preferences of individuals[46]. In conjoint exercises, respondents are asked to make choices between profiles consisting of different levels of attributes. The relative value or importance of these attributes can then be quantified[47–49]. The term *conjoint* means that multiple attributes can be "considered jointly"[50]. One reported advantage of conjoint analysis in implementation science is that stakeholders are required to state their preference for precise attributes of strategies which, in turn, can enhance the rigor, preciseness and contextual relevance of strategies seeking to increase the uptake of best practices[46]. In implementation research, the use of conjoint analysis is just beginning. For example, a study by Waltz et al. used this method to develop expert recommendations for strategies that can best support the implementation of three clinical practice changes for veteran mental health[51]. Another group of researchers used discrete choice experiment, a type of conjoint analysis, to estimate the influence of 16 attributes of mental health practice changes (contextual and social, content and practice change process attributes) on the preferences of educators[52]. More recently, Williams et al. determined the most preferred implementation strategies to support new psychosocial practices among clinicians, clinical supervisors, and administrators[53]. Conjoint analysis has also been used to estimate preferences for clinical practice guidelines features[54].

In discrete-choice experiment, a widely used derivative of conjoint analysis, respondents are presented with two or more sets of hypothetical profiles incorporating multiple attributes of interest and are asked to select their preferred profile[49]. Respondents must consider the trade-offs

---

[m] In this introduction, *attributes* is used interchangeably with *factors* for the purposes of describing the weight elicitation methods used across multiple fields. *Attributes* is used in the economics literature.

across a set of attributes which in turn reflect the value that they attach to those attributes. These choices allow for an efficient and highly accurate estimation of the individual weights for each dimension[55,56]. Weights are numbers representing the attractiveness of each dimension-level. Further information can be estimated from analysis of discrete-choice experiments such as dimension importance, in other words, the relative influence that each overall dimension has in the model.

The study reported in this paper was conducted in the context of the development of a brief multidimensional index called the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The PIRE-CDMI includes five items, each one representing a key EBP domain that was identified based on a comprehensive item development and reduction process including a literature review, nominal group technique, Rasch analysis[18], expert consensus, and qualitative revision and rewriting as described in chapters 5 and 7 of this dissertation. The most recent version of the PIRE-CDMI was used as the basis for the weight elicitation exercise in this study; the five dimensions will take on the same role as the attributes described in weight elicitation tasks.

The five dimensions included in the PIRE-CDMI are: (1) a clinician's self-reported use of research evidence when faced with a practice uncertainty; (2) a clinician's self-efficacy (also known as confidence) in applying research evidence to practice; (3) the availability of resources to help the clinician integrate research evidence into practice; (4) the effort required by the clinician to integrate research evidence into practice; and (5) the frequency with which a clinician keeps up to date with research evidence. The conceptual model for this measure is formative, that is, dimensions form the construct in an additive way and having higher levels in each dimension is likely to increase engagement in EBP. However, evidence is needed as to the relative contribution of each dimension to a clinician's overall propensity to integrate research evidence into clinical decision-making (i.e., the total score).

Objectives

The global aim of the study is to develop a scoring algorithm for the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). The specific objective was to estimate the relative importance of five key dimensions to the propensity to integrate research evidence into clinical decision-making, information that will be used in the scoring algorithm. The secondary objective was to contribute evidence of the ability of the PIRE-

CDMI to distinguish between known groups of OTs and PTs and to compare with a global self-rating of EBP.

Methods

Study sample

The target population for this study was comprised of: (1) OTs and PTs who were licensed and working in Canada for at least one year; (2) clinical managers of OTs, PTs or both working in Canada; (3) experts in EBP in rehabilitation characterized as (a) self-proclaimed knowledge in EBP in rehabilitation and (b) experience conducting research in EBP or having at least one EBP-related publication. Participants were eligible if they spoke English or French and had access to a computer to complete the online survey. Participants were not excluded based on setting or clinical context.

Rehabilitation clinicians and clinical managers were recruited using convenience sampling through ads posted on the social media of the research team (Twitter and Facebook), in national professional associations listservs (Canadian Physiotherapy Association and the Canadian Association of Occupational Therapy) and in the McGill University health center network. Experts in EBP were first recruited purposively from a list of potential participants provided by the research team, followed by snowball sampling to increase the pool. Potential participants were invited to click on the survey link for more information on the study and start the survey after attesting to being eligible and consenting. All participants who completed the survey received an electronic $5 coffee shop gift card (except for those that opted out).

Based on the estimation exercise conducted before data collection (whereby the choice exercise design was tested using Orme's Sawtooth Software[57]) our targeted sample size was 60 to produce precise utility estimates with an alpha level of 0.05 and standard errors less than or equal to 0.05 given the parameters of our best-worst scaling (BWS) design (described below).

Data collection

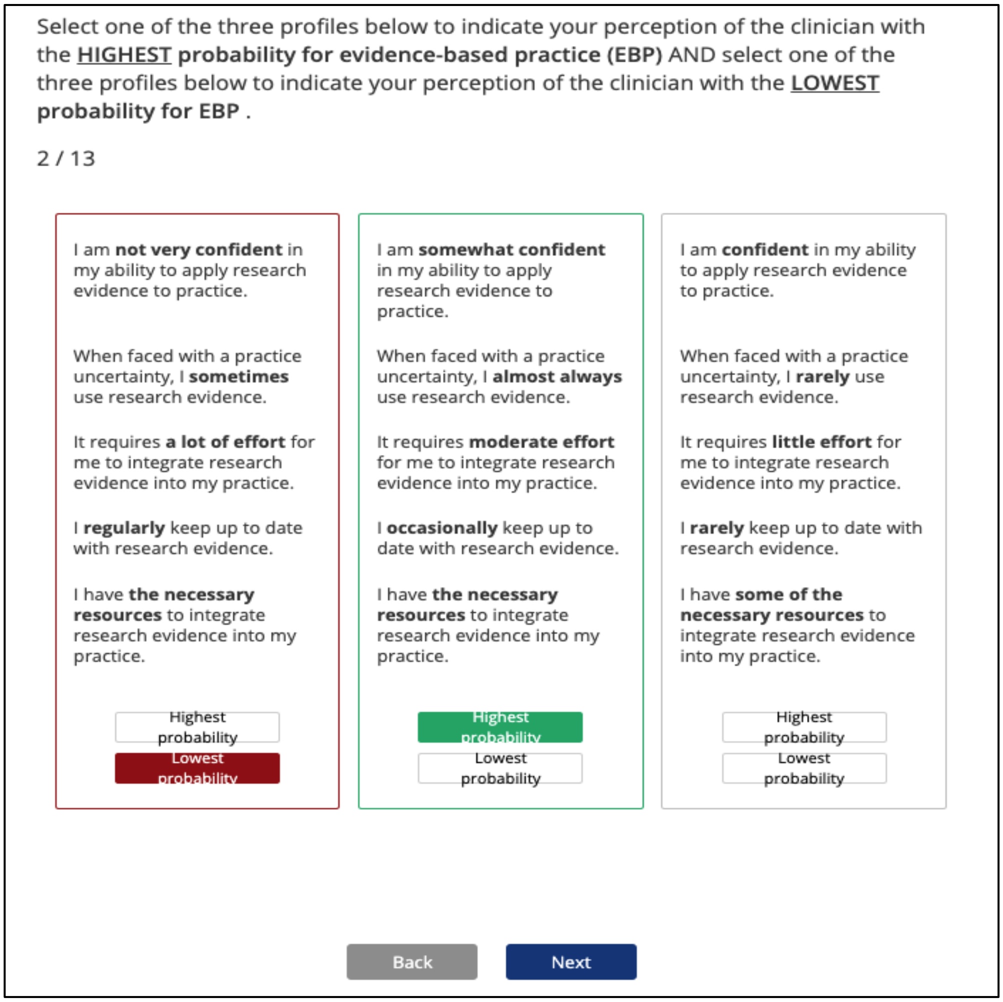
Data were collected through an online survey using Sawtooth software (active from Feb 17, 2022, to March 30, 2022) which mounted the weight elicitation exercise. Conjoint analysis was selected as the weight elicitation exercise for this study because it involves explicit trade-offs between EBP dimensions which are reflective of authentic practice contexts in that these individual and contextual factors do not occur independently of one another. Other methods to estimate weights such as rating EBP profiles on a numeric scale were not applicable to the context of study because providing a quantitative estimate or price (as in the willingness-to-pay method) of an overall EBP profile was deemed to be arbitrary.

Specifically, this study employed a methodological extension of discrete-choice experiment called *best-worst scaling* (BWS), also known as maximum-difference scaling[58]. The *multi-profile case* variant of BWS entails a dual selection of profiles whereby respondents identify the most preferred (i.e., best) and least preferred (i.e., worst) profiles instead of just selecting one profile per set [59]. BWS is a means of collecting more information and reducing standard error, without increasing respondent fatigue[59,60]. With BWS, fewer participants are generally required to estimate precise preferences suggesting greater statistical efficiency compared to conventional discrete-choice experiment[55]. This exercise can be seen as ordinal in nature and leverages the person's ability to identify profiles more easily at extremes[61].

The survey was piloted with five clinician-researchers in rehabilitation to ensure clarity of instructions and profiles, appropriateness and understanding of BWS tasks and functioning of the survey. Pertinent feedback was integrated (e.g., bolding the lettering of the levels to help respondents quickly visually discriminate between profiles). From the pilot experience, the estimated completion time for the survey was 10-15 minutes, deemed feasible for the main study. The survey consisted of (1) study information and consent; (2) a rationality test to ensure that respondents had become familiar with the dimensions and understood the task; (3) the 12 rounds of the BWS task; (4) four questions regarding acceptability of the exercise; and (5) demographic questions. Clinicians specifically started the survey by completing the five multiple-choice PIRE-CDMI items and a global self-rating of EBP (0-10 scale) as it applied to their own clinical practice.

As illustrated in Figure 1, all possible PIRE-CDMI dimensions and levels were used in creating profiles for the BWS task. Each task (or page) included three profiles each with varying levels of the five dimensions (three possible levels per dimension). Table 1 presents a comprehensive list of the dimensions and levels used in the survey. Respondents were asked to select which theoretical profile, out of the three profiles, represented a clinician with the *highest* probability for EBP and one with the *lowest* probability for EBP; the specific wording of the choice question is included in Figure 1. Thus, out of the three profiles in one task, respondents selected two - one highest and one lowest. There was no “none” or “opt-out” option in the task.

Figure 1. Example of the best-worst scaling task



The process comprised 12 rounds of BWS tasks. 12 rounds were selected to limit the cognitive burden on participants whilst still generating precise estimates[62]. A fractional factorial and balanced overlap design was used within the choice-based conjoint-analysis package provided by Sawtooth Software. This design estimates main effects and two-way interactions and organizes the levels to be shown in equal numbers. Each respondent received a randomized version of a subset of the full-choice design whereby each profile is assumed to be a well-balanced and near-orthogonal fraction of the full-choice design[63]. This randomized design approach minimized the number of rounds of BWS per respondent while still collecting data on various subsets[48]. Respondents indicated the level of difficulty of the exercise, comprehensibility of profiles, and degree of frustration with the exercise, each with a 4-point Likert scale. They also reported whether they would be willing to participate in a similar study in the future (yes/no).

Table 1: Dimensions, statements, and response levels used in the best-worst scaling exercise

Dimension	Statement	Response level
Use of research evidence	When faced with a practice uncertainty*, I ____ research evidence. *uncertainty: a situation in which there is a gap in your knowledge relating to a clinical decision	almost always use
		sometimes use
		rarely use
Self-efficacy	I am ____ in my ability to apply research evidence to practice.	confident
		somewhat confident
		not very confident
Resources	I have _____ to integrate research evidence into my practice. *Examples: paid time to consult the research evidence, access to journals, access to a computer, access to necessary therapeutic material...	the necessary resources*
		some of the necessary resources*
		few of the necessary resources*
Effort	It requires_____ for me to integrate research evidence into my practice.	little effort
		moderate effort
		a lot of effort
Keep up to date	I _____ keep up to date with research evidence.	regularly
		occasionally
		rarely

Statistical analysis

For analysis, participants were excluded when they (1) did not complete the survey, (2) did not pass the rationality test, and (3) were flagged as having answered irrationally or randomly, defined as having a root-likelihood (RLH) fit statistic of lower than 0.4.

Part-worth utilities and average dimension importance were estimated for the entire sample, both languages combined, using the hierarchical Bayesian analytic tool from Sawtooth

Software (Lighthouse studio version 9.13.2). Part-worth utilities and importance were modeled using a standard utility function whereby part-worth attributes were specified[64]. Hierarchical Bayes allows for the modeling of non-linear relationships and the estimation of individual-level weights which is considered an advantage as people may value attributes differently. Hierarchical Bayesian analysis uses an algorithm whereby individual respondent models (also called "lower-level models" or "posterior weights" and governed by a multinomial logit model) are balanced with the sample's mean parameters (also called "upper-level models" or "prior weights" and described by a multivariate normal distribution)[65]. Specifically, the algorithm first estimates how different each individual's weights are from the other respondents. Then, the algorithm adjusts each individual's weights so that there is an optimal degree of fit between the individual's choices and the sample average[66]. The optimal degree is determined by the amount of data and the consistency of the data that each individual provides. This algorithm is run 10,000 times (whereby the sample average is continuously updated) before converging and producing stable estimates of weights[65]. The relative importance of each dimension was calculated by subtracting the lowest utility level from the highest utility level and dividing that score by the total dimension range.

Differences in the ranking of average importance of dimensions by professional group were tested using chi-square tests. Descriptive statistics were used to tabulate the data on the feasibility and acceptability of the BWS method. To develop a scoring algorithm for the PIRE-CDMI, the part-worth utilities were recalibrated to generate a total PIRE-CDMI score from 0-100. To simplify future use of the index, simple part-worth utilities were also calculated for easier scoring by hand. Each clinician was attributed a total score through a simple additive formula of the part-worth utilities for each item-level response to the PIRE-CDMI. The relationship between index scores and profession (OT, PT) were tested using simple linear regression. It was hypothesized that PTs would score at least a minimally important difference higher than OTs due to pre-existing data from the literature[10,18,67,68] and from chapter 5 of this dissertation suggesting that PTs tend to score higher on EBP scales. PIRE-CDMI item-item correlations (polychoric) and the correlation between PIRE-CDMI scores and the global rating of EBP (0-100) (Spearman) were calculated. We hypothesized that inter-item correlations would be low to moderate and correlation with the global rating would be moderate. Correlation coefficients (r) of $\geqslant 0.8$ were considered strong, $0.4 \leqslant r < 0.8$ moderate, and $r < 0.4$ as low[69]. All statistical analyses were conducted using SPSS (version 28.0.1.0). Ethics approval was obtained from *The Faculty of Medicine and Health Sciences Institutional Review Board* at McGill University for this study before commencement and recruitment.

Results

Sample

Eighty-six participants accessed the survey and 66 (77%) completed it. Among those who completed the survey, 61 passed the rationality test (71%). The sample consisted of 15 occupational therapists (25%), 13 physical therapists (21%), 16 clinical managers (26%), and 17 experts in EBP (28%). Fifty-two (85%) identified as female, 8 as male (13%) and one preferred not to respond (2%). The mean age of participants was 42 years old (SD 10.8) ranging from 22 to 67 years. More than two thirds of participants were working in Quebec at the time of the survey (42, 69%) while others worked in Ontario (12, 20%), Alberta (4, 7%), Manitoba (2, 3%) and New Brunswick (1, 1%). Table 2 presents more demographic information.

Table 2: Descriptive demographics of the sample of clinicians (n=28), clinical managers (n=16) and experts in EBP (n=17)

	Mean (SD) or n (%)	
Clinicians (OTs, PTs) (n=28)		
Years of practice	11.3	8.7
Full-time schedule	19	66
Clinical setting affiliated with a university	10	36
Clinical population		
Orthopaedics, musculoskeletal	12	43
Pediatric	9	32
Neurology	3	11
Geriatrics	2	7
Acute care	1	4
Mental health	1	4
Clinical settings		
Private practice	15	54
Rehabilitation center	6	21
General hospital – acute care	3	11
Home visiting agency	3	11
Primary health care	2	7

Other	2	7
Long-term care/complex continuing care	1	4
Community agency	1	4
General hospital – long-term care	1	4
Clinical managers (n=16)		
Years of experience as a manager	9.4	8.7
Professionals supervised by managers		
OTs	3	19
PTs	3	19
Both	10	62
Full-time schedule	13	81
Clinical setting affiliated with a university	10	63
Clinical population		
Orthopaedics, musculoskeletal	5	31
Varied	4	25
Pediatrics	3	19
Adults - unspecified	2	13
Other	2	13
Clinical settings		
Rehabilitation center	6	38
General hospital – acute care	6	38
Private practice	3	19
Home visiting agency	2	13
General hospital – long-term care	1	6
Long-term care/complex continuing care	1	6
Primary health care	1	6
Experts in EBP (n=17)		
Profiles		
University researcher or professor	11	65
PhD student or post-doc	1	6
Professional rehab association, order or regulatory body	3	18
Consultant or clinical specialist	2	12
Focus of EBP research/work		
OTs	7	41
PTs	8	47
Both	2	12
Settings where EBP expertise was developed		
General hospital – acute care	10	59
Rehabilitation center	7	41
University	5	29
Private practice	3	18
Community agency	3	18
Regulatory	2	12

Primary health care	1	6
Home visiting agency	1	6
Outpatient hospital	1	6

Continuous variables are presented as means and standard deviation (SD) while categorical variables are presented as frequencies

OT, occupational therapist; PT, physical therapist; EBP, evidence-based practice

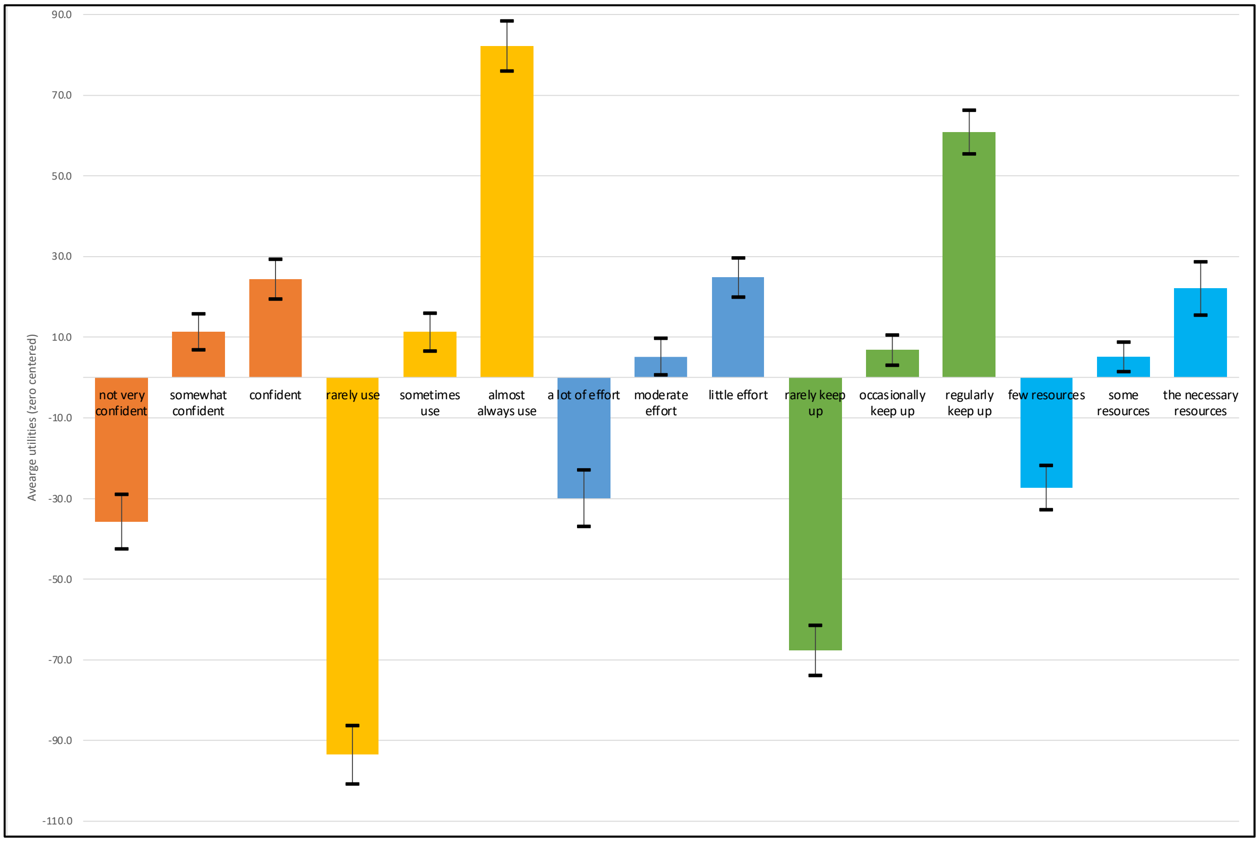
Estimated utilities

The part-worth utilities for the entire sample are reported in Table 3. Utilities in this context refers to how much each dimension-level builds a clinician's propensity of integrating research evidence into clinical decision-making. Figure 2 illustrates the range in part-worth utilities per item per level for the entire sample. The 95% confidence intervals for the part-worth utilities did not overlap between levels of the same dimension (as seen in Figure 2). Standard errors for the estimated utilities are 0.05-0.06 for each dimension (except for "rarely use", 0.07). The mean root likelihood (RLH) fit statistics for the utilities of individual respondents was 0.72 out of 1 (minimum: 0.48, maximum: 0.87) which indicates how well a choice models fits the data set wherein higher values denote better fit[65].

Table 3. Part-worth utilities per dimension level for the sample (n=61)

Dimension levels	Part-worth utilities	Standard deviation	Lower 95% CI	Upper 95% CI
not very confident	-35.8	27.0	-42.5	-29.0
somewhat confident	11.4	17.7	6.9	15.8
confident	24.4	19.9	19.4	29.4
rarely use	-93.5	28.8	-100.7	-86.3
sometimes use	11.3	18.7	6.6	16.0
almost always use	82.2	24.7	76.0	88.4
a lot of effort	-29.9	27.8	-36.9	-23.0
moderate effort	5.1	18.1	0.6	9.7
little effort	24.8	19.5	19.9	29.7
rarely keep up	-67.7	24.8	-73.9	-61.4
occasionally keep up	6.8	14.9	3.1	10.6
regularly keep up	60.8	21.6	55.4	66.3
few resources	-27.3	22	-32.8	-21.8
some resources	5.2	14.5	1.5	8.8
the necessary resources	22.1	26.3	15.5	28.8

Figure 2. Histogram illustrating the part-worth utilities (zero-centered) per dimension with 95% confidence intervals



Relative importance

Table 4 presents the average relative importance of each EBP dimension for the total sample and per professional group. Chi-square statistics do not suggest an important difference in ranking by professional group (X^2 (df 12, N=61) =7.6, p=0.82).

Table 4. Average relative importance for each dimension included in the index (n=61)

Dimension	%
Use of research evidence	35.2
Keep up to date	25.8
Self-efficacy	13.6
Effort	13.1
Resources	12.4

Feasibility and acceptability of the best-worst scaling method

Table 5 summarizes the feasibility and acceptability results for the BWS exercise across the sample.

Table 5. Feasibility and acceptability indicators of the best-worst scaling method for EBP profiles with clinicians, managers, and experts (n=61)

	n	**%**
Difficulty of the exercise		
1 - not difficult at all	3	4.9
2- a bit difficult	11	18.0
3- difficult	32	52.5
4 - very difficult	15	24.6
Comprehensibility of the profiles		
1- very comprehensible	15	24.6
2- comprehensible	28	45.9
3- incomprehensible	12	19.7
4- completely incomprehensible	6	9.8
Degree of frustration with exercise		
1- not at all frustrated	10	16.4
2- a bit frustrated	23	37.7
3- frustrated	20	32.8
4- very frustrated	8	13.1
Willingness to participate again		
Yes	46	75.4
No	15	24.6
Completion rate of survey		
Complete	66	76.7
Incomplete	20	23.3
Time taken to complete the entire survey		
Mean (SD)	25.7 min	(28.8 min)
Median	17.6 min	

PIRE-CDMI scoring algorithm

Table 6 presents the PIRE-CDMI scoring algorithm based on recalibrated part-worth utilities and simple part-worth utilities for the entire sample. Table 7 presents descriptive information and evidence that the PIRE-CDMI scores are interpretable as hypothesized. Appendix I presents the inter-item correlation matrix for the PIRE-CDMI. Regression suggests that there is an important difference between OTs and PTs on the index whereby, on average, PTs score 13 points higher (95% CI 0.6 -25.3) than OTs. The mean self-reported global rating of EBP was 7.7 out of 10 (95% CI 7.0-8.4) with a Shapiro-wilk estimate suggesting a departure from normality (W=0.91, p=0.03). The Spearman's rank-order correlation demonstrated a positive correlation between PIRE-CDMI scores and global ratings of EBP ($r_s(26) = 0.56$, $p = .002$).

Table 6. Scoring algorithm for the PIRE-CDMI

	Full part-worth utilities **(range 0-100)**	**Simple** part-worth utilities* **(range 0-46)**
Use of research evidence		
rarely use	0.0	0
sometimes use	22.4	6
almost always use	37.5	10
Self-efficacy		
not very confident	0.0	0
somewhat confident	10.1	3
confident	12.8	4
Resources		
few resources	0.0	0
some resources	6.9	2
the necessary resources	10.6	3
Effort		
a lot of effort	0.0	0
moderate effort	7.5	2
little effort	11.7	4
Keep up to date		
rarely keep up	0.0	0
occasionally keep up	15.9	4
regularly keep up	27.4	8
*Simple part-worth utilities were calculated to ease the process of scoring by hand and were based on linear transformations relative to the lowest actual weight (6.9).		

Table 7. Descriptive information and evidence for interpretability of PIRE-CDMI (n=28)

	Clinicians (n=28)	**OT (n=15)**	**PT (n=13)**
Mean total score (SD)	72.5 (16.9)	66.4 (16.4)	79.4 (15.2)

Median	76	62.8	84.6
Minimum	40	40	48.4
Maximum	100	96.3	100
Interquartile range	29.7	26.3	20.8
Skewness	-0.23	0.20	-0.84
Kurtosis	-1.15	-1.01	-0.02
Shapiro-Wilk test of normality, df 28	0.948 (p=0.18)		
Floor effect (n, %)	0 (0%)		
Ceiling effect (n, %)	1 (3.6%)		
Mean completion time of 5 items included in the PIRE-CDMI (SD)	1.2 minutes (0.7 minutes)		
Minimum completion time	0.4 minutes		
Maximum completion time	3.1 minutes		
PIRE-CDMI scores were calculated using the full part-worth utilities. SD, standard deviation; OT, occupational therapist; PT, physical therapist; df, degrees of freedom			

Discussion

The global aim of this study was to develop a scoring algorithm for a brief, multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into clinical decision-making called the PIRE-CDMI. This is the first study to use choice-based experiments, specifically BWS, to estimate the relative importance of five EBP dimensions from multiple stakeholders, information that was integrated into a scoring algorithm.

Estimated utilities

The multi-profile BWS approach allowed for the efficient generation part-worth utilities for each dimension-level of the PIRE-CDMI from the perspectives of OTs, PTs, clinical managers, and experts in EBP. As expected, the part-worth utilities for dimension-levels were ordered for all five items of the PIRE-CDMI, that is, the lower response options had the least weight, and the highest response option had the most weight. Interestingly, for the *self-efficacy* item, the middle and high response levels demonstrated a smaller difference in weights compared to the third response level; this indicates that participants found "somewhat confident" and "confident" to be similar quantitatively. Part-worth utilities must be interpreted in a relative sense as the lowest response option does not necessarily indicate undesirable, but relatively worse than the other two levels in the same dimension. As previously described by authors who studied propensity for practical participatory evaluation[70], having an index related to *propensity* for EBP as opposed to an index of barriers to EBP was, in part, to avoid the negative connotation that individuals are resistant to change and a self-realizing prophecy[71]. As such, the focus is shifted from what does not work to what works and can be improved. When interpreting results, future PIRE-CDMI respondents will score on a continuum from 0 or neutral propensity to 100 or high propensity; a score of 0 does not indicate no propensity to integrate research evidence, but neutral propensity which builds with each item level-increment. This way, the index builds on strengths rather than weaknesses and clinicians may perceive the PIRE-CDMI as a constructive and helpful tool rather than an ominous one.

The two dimensions viewed as most impactful on propensity to integrate research evidence into clinical decision-making were *use of research evidence* and *keep up to date*. The predominance of the *use of research evidence* item (35% average relative importance) is not surprising given that this item reflects the behavior of interest, a clinician's self-reported use of research evidence when faced with a clinical uncertainty. Of the 46 points (based on the simple part-worth utilities), 10 points are allocated to the *use of research evidence* item whereby a six-point difference exists between "rarely use" and "sometimes use", and a four-point difference exists between "sometimes use" and "almost always use". This signals that the middle response option is not exactly in the middle and reinforces the importance of estimating part-worth utilities to establish a true score.

Participants highly valued the *keep up to date* item (26% average relative importance) which speaks to the importance of maintaining periodic research evidence-seeking behaviors. This has implications for the development of future interventions which could prioritize evidence-based activities such as keeping up with research evidence over other factors which are rated lower in average importance. As elucidated in chapter 7, this item intentionally does not specify the source of research evidence. Indeed, clinicians report gaining research evidence from a broad selection of sources including peer consultation, journal clubs and email alerts[72–77] . Conversely, *self-efficacy, effort* and *resources* were viewed as much less impactful on propensity than the two other dimensions. Surprisingly, the *resources* item scored lowest on average relative importance (12%) albeit very similar to the *effort* (13%) and *self-efficacy* (14%) items. Given what is known on the influence of organizational resources on EBP[5,11–13,28,75,78,79], a total of 3 of 46 points appears very small. A possible explanation may be the new contextualized interpretation of this factor wherein the perceived impact of resources is offset when combined with the four other factors. This inconsistency may also be due to the narrower range in response options associated with the *resources* item (“few resources”, “some resources”, “the necessary resources”) which may not have the same resonance as items with stronger descriptors such as, “rarely use” and “almost always use” (*use of research evidence* item) and “rarely” and “regularly” (*keep up to date* item).

Fourty-seven participants (77%) perceived the choice exercise to be difficult or very difficult, and five individuals failed the rationality test (7%). The complexity of the BWS exercise is noteworthy but unsurprising given the granularity of profile comparisons and abstract thinking required to choose profiles. As expressed by Devlin et al.[80], participating in valuation exercises elicits all four major classes of cognitive functions as proposed by Lezak[81] including receptive functions, memory and learning, thinking, and expressive functions. This important cognitive effort may explain why 43 participants (70%) reported being a bit frustrated or frustrated with the BWS exercise and why 23% did not complete the survey. Eighteen participants (30%) rated the profiles as being not comprehensible which may have been influenced by certain illogical profile configurations (e.g., having the lowest response level on all options except the *use of research evidence* item which had the highest level). Profile configuration constraints were not introduced into the design of profiles to avoid masking important findings. Nevertheless, 46 respondents (75%) reported that they would be willing to participate again.

Integration of the utilities into the PIRE-CDMI scoring algorithm

Integrating the estimated part-worth utilities into a scoring algorithm allowed for the development of a total PIRE-CDMI score that is consistent with how stakeholders perceive the relative value of each dimension-level by way of their profile choices. PIRE-CDMI scores were normally distributed as suggested by the Shapiro-Wilk test and visual examination of the histogram and QQ-plot. No respondent scored lower than 40 and only one respondent scored the maximum (100) which is insufficient to indicate ceiling effects. In accordance with the hypothesis, PTs scored 13 points higher than OTs on the index, which is larger than the estimate of a minimally important difference (8.5 points)[82] . The PIRE-CDMI shows a moderate positive correlation (0.6) with the global self-ratings of EBP which is a reasonably expected relationship given that the PIRE-CDMI is multidimensional (i.e., measures different factors) while the global rating is unidimensional. It is also possible that the wording of the global self-rating question which may have been difficult for clinicians to answer, (“On the 0-10 scale; please rate your own overall likelihood to integrate evidence into practice"). Without contextual or temporal cues related to the integration of evidence into practice, clinicians may have found the question vague and abstract. The PIRE-CDMI inter-item polychoric correlations were between 0.02 and 0.68 which supports our hypothesis of low to moderate correlations for a multidimensional measure. This confirms that dimensions represent different facets of the construct of propensity to integrate research evidence into CDM. Internal consistency was not evaluated and would not be appropriate for a multidimensional index, as analyses such as Cronbach's alpha seek to verify the unidimensionality between items of the same scale[83,84].

Respondents took, on average, one minute to complete the PIRE-CDMI, unlike the 15 minutes to complete the 70-item survey from which the PIRE-CDMI was developed[18]. This short, easy to administer questionnaire may be more feasible for busy clinicians and may improve completion rates in the future. Further, the PIRE-CDMI fulfills most of the criteria for pragmatic measures as it is low in burden for respondents and administrators, broadly applicable, actionable, and unlikely to cause harm; still, future field testing must be conducted to fulfill the criteria[85].

Strengths and limitations

In terms of limitations, the large standard deviations of the estimated part-worth utilities denote an important variability in values across the sample which was expected given the heterogeneity of respondents. However, the 95% confidence intervals of part-worth utilities are narrow considering the wide range in values and the lack of overlap between level-item confidence intervals suggest acceptable precision of average part-worth utilities. The standard errors of the aggregate logit between 0.05 and 0.06 represent estimation uncertainty and are sufficiently low to rule out poor precision on estimates. In designing the choice task, authors aimed to reduce random variability and optimize the precision of estimates by decreasing the number of choice tasks to 12, integrating a rationality test and eliminating those who failed said test. Still, the complexity of choice-based exercises and the cumulative cognitive burden has been shown to negatively impact estimation of part-worth utilities and increase variance[47,86]. Each set comprised three full profiles with five dimensions per profile. Given acceptability results, it is plausible that this may have been too much information to process within one's working memory as adults can only process about seven pieces of information at once[87]. This may have led respondents to use simplified decision-making heuristics such as only relying on specific dimensions to govern choices[88]. Future studies examining EBP factors could consider using two full profiles (i.e., a typical discrete-choice experiment method) to limit the cognitive burden. However, a two-profile set would necessitate a much larger sample size and a higher number of choice tasks which could increase respondent fatigue and incompletion rates.

Additionally, respondents assessed a hypothetical profile of a rehabilitation clinician; real-world assessment of clinician behavior may produce different results due to other EBP dimensions or confounders unaccounted for in our study. It is also possible that a different order of dimensions may have produced different patterns of dimension importance and choosing different levels or item wording may have produced different part-worth utilities[61]. If the PIRE-CDMI is modified in the future, items could be integrated into another choice-based valuation exercise to update the estimated weights. The findings on average relative importance must be interpreted with caution because the results are directly affected by the range of levels for each dimension and the number of dimensions. There is the possibility of having attribute bias which is the systematic preference of one or more dimensions in a choice-based activity[61]. This may have occurred with the *use of research evidence* item due to its broad nature. However, given that the *keep up to date* item generated similarly high valuations, it is unlikely that the presence of the *use of research evidence* was overbearing. The estimated weights are closely linked to the PIRE-CDMI structure which was developed with and for OTs and PTs in Canada; the generalizability of the weighted PIRE-CDMI to other professions would require extensive adaptation and testing.

Although the use of a motivated sample of clinicians may have slightly skewed PIRE-CDMI scores towards higher propensity, it is improbable that selection bias has threatened the internal validity of the estimated part-worth utilities. It is likely that non-respondents would have made choices on the BWS tasks similarly to respondents. Among clinicians and clinical managers, our sample includes a broad representation of clinical settings, patient populations and other demographics. While there was a considerable representation of respondents working in Quebec (70%), the healthcare landscape in Quebec resembles other provinces (with the exception of the territories) owing to recent reorganizations towards centralized governments and a similar health spending per person[89].

Applications in practice and future research

The PIRE-CDMI may be used as a clinical outcome measure with OTs and PTs to measure propensity to integrate research evidence into clinical decision-making. For instance, the index can be used in effectiveness studies of interventions focused on the implementation of EBPs with rehabilitation professionals, however, further field testing is necessary to substantiate the use of the index as an outcome measure in clinical research. These studies could estimate the (1) relationship between PIRE-CDMI scores and comprehensive measures of EBP; (2) relationship between PIRE-CDMI scores and other indicators relating to the integration of research evidence into CDM (e.g., frequency of consulting scientific journals); (3) sensitivity to change after an EBP intervention; and (4) associations with variables such as time since graduation, profession, clinical setting and area of practice, workload, and peer isolation.

Although the PIRE-CDMI is likely to be useful for other stakeholders including clinical managers and regulators, this index was not developed for regulatory or jurisprudence purposes. If used for purposes outside of clinical research, there is a high likelihood of social-desirability bias as clinicians would feel obligated to respond in a desirable manner to avoid social consequences. Future research is necessary to substantiate the use of this index as a self-reflection tool for clinicians or a discussion guide for clinical managers.

As there is a possibility of making mistakes when scoring by hand, it is important for future users of the PIRE-CDMI to consider the approach to scoring such as computerized scoring which decreases scoring miscalculations but increases complexity and cost[69].

Conclusion

Various individual and organizational factors influence a rehabilitation clinician's propensity to integrate research evidence into the clinical decision-making process. The interdependent nature of these influencing factors should be considered when measuring EBP. This study describes the elaboration of a weighted scoring algorithm for a multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into the clinical decision-making process. Best-worst scaling, a choice-based exercise, was used to determine the quantitative relative weighting of five key dimensions influencing EBP propensity. The dimensions *use of research evidence* and *keep up to date* had higher relative importance compared to the other three dimensions, *self-efficacy*, *effort,* and *resources*. The estimated weights were integrated into a scoring algorithm allowing for a single propensity score to be representative of the gains in one domain against the losses in another domain. This study also contributes validity evidence regarding internal structure, relationships to other variables and feasibility which support the use of PIRE-CDMI results to draw inferences about OTs and PTs.

References

1. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *BMJ Evid-Based Med*. 2002;7(2):36-38.

2. Haynes RB, Sackett DL, Richardson WS, Rosenberg W, Langley GR. Evidence-based medicine: How to practice & teach EBM. *Can Med Assoc J*. 1997;157(6):788.

3. CAOT. Profile of occupational therapist practice in Canada. Published online 2012. Accessed September 17, 2020. https://www.caot.ca/document/3653/2012otprofile.pdf

4. Group NPA. Competency profile for physiotherapists in Canada. *Natl Physiother Advis Group*. Published online 2017. https://www.peac-aepc.ca/pdfs/Resources/Competency%20Profiles/Competency%20Profile%20for%20PTs%202017%20EN.pdf

5. da Silva T, da Cunha Menezes Costa L, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. 2015;20(3):388-401. doi:10.1016/j.math.2014.10.009

6. Graham F, Robertson L, Anderson J. N ew Z ealand occupational therapists' views on evidence-based practice: A replicated survey of attitudes, confidence and behaviours. *Aust Occup Ther J*. 2013;60(2):120-128.

7. Bernhardsson S, Johansson K, Nilsen P, Öberg B, Larsson MEH. Determinants of Guideline Use in Primary Care Physical Therapy: A Cross-Sectional Survey of Attitudes, Knowledge, and Behavior. *Phys Ther*. 2014;94(3):343-354. doi:10.2522/ptj.20130147

8. Upton D, Stephens D, Williams B, Scurlock-Evans L. Occupational Therapists' Attitudes, Knowledge, and Implementation of Evidence-Based Practice: A Systematic Review of Published Research. *Br J Occup Ther*. 2014;77:24. doi:10.4276/030802214X13887685335544

9. Saunders H, Gallagher-Ford L, Kvist T, Vehviläinen-Julkunen K. Practicing Healthcare Professionals' Evidence-Based Practice Competencies: An Overview of Systematic Reviews. *Worldviews Evid Based Nurs*. 2019;16(3):176-185. doi:https://doi.org/10.1111/wvn.12363

10. Thomas A, Zoubi FA, Mayo NE, et al. Individual and organizational factors associated with evidence-based practice among physical and occupational therapy recent graduates: A cross-sectional national study. *J Eval Clin Pract*. 2020;n/a(n/a). doi:https://doi.org/10.1111/jep.13518

11. Scurlock-Evans L, Upton P, Upton D. Evidence-Based Practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014;100(3):208-219. doi:10.1016/j.physio.2014.03.001

12. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to Evidence-Based Practice Implementation in Physiotherapy: A Systematic Review and Meta-Analysis. *Int J Qual Health Care*. 2021;(mzab093). doi:10.1093/intqhc/mzab093

13. Juckett LA, Wengerd LR, Faieta J, Griffin CE. Evidence-Based Practice Implementation in Stroke Rehabilitation: A Scoping Review of Barriers and Facilitators. *Am J Occup Ther Off Publ Am Occup Ther Assoc*. 2020;74(1):7401205050p1-7401205050p14. doi:10.5014/ajot.2020.035485

14. Mayo N. *ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*. 1st edition. International Society for Quality of Life Research; 2015.

15. Buchanan H, Siegfried N, Jelsma J. Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review: Evidence-based Practice Survey Instruments. *Occup Ther Int*. 2016;23(2):59-90. doi:10.1002/oti.1398

16. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gómez JE. Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the

literature: Evidence-based physiotherapy instruments. *J Eval Clin Pract*. 2014;20(6):767-778. doi:10.1111/jep.12180

17. Fernández-Domínguez JC, de Pedro-Gómez JE, Morales-Asencio JM, Bennasar-Veny M, Sastre-Fullana P, Sesé-Abad A. Health Sciences-Evidence Based Practice questionnaire (HS-EBP) for measuring transprofessional evidence-based practice: Creation, development and psychometric validation. Gagnier JJ, ed. *PLOS ONE*. 2017;12(5):e0177172. doi:10.1371/journal.pone.0177172

18. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

19. Hadley J, Hassan I, Khan KS. Knowledge and beliefs concerning evidence-based practice amongst complementary and alternative medicine health care practitioners and allied health care professionals: a questionnaire survey. *BMC Complement Altern Med*. 2008;8(1):1-7.

20. Heiwe S, Kajermo KN, Tyni-Lenné R, et al. Evidence-based practice: attitudes, knowledge and behaviour among allied health care professionals. *Int J Qual Health Care*. 2011;23(2):198-209. doi:10.1093/intqhc/mzq083

21. Iles R, Davidson M. Evidence based practice: a survey of physiotherapists' current practice. *Physiother Res Int*. 2006;11(2):93-103.

22. Jette DU, Bacon K, Batty C, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. *Phys Ther*. 2003;83(9):786-805.

23. Lyons C, Casey J, Brown T, Tseng M, McDonald R. Research knowledge, attitudes, practices and barriers among paediatric occupational therapists in the United Kingdom. *Br J Occup Ther*. 2010;73(5):200-209.

24. MacEwan Dysart A, Tomlin GS. Factors Related to Evidence-Based Practice Among U.S. Occupational Therapy Clinicians. *Am J Occup Ther*. 2002;56(3):275-284. doi:10.5014/ajot.56.3.275

25. McCluskey A. Occupational therapists report a low level of knowledge, skill and involvement in evidence-based practice. *Aust Occup Ther J*. 2003;50(1):3-12. doi:10.1046/j.1440-1630.2003.00303.x

26. Nilsagård Y, Lohse G. Evidence-based physiotherapy: A survey of knowledge, behaviour, attitudes and prerequisites. *Adv Physiother*. 2010;12:179-186. doi:10.3109/14038196.2010.503812

27. Philibert DB, Snyder P, Judd D, Windsor MM. Practitioners' reading patterns, attitudes, and use of research reported in occupational therapy journals. *Am J Occup Ther*. 2003;57(4):450-458.

28. Salbach NM, Guilcher SJ, Jaglal SB, Davis DA. Determinants of research use in clinical decision making among physical therapists providing services post-stroke: a cross-sectional study. *Implement Sci*. 2010;5(1):77. doi:10.1186/1748-5908-5-77

29. Salls J, Dolhi C, Silverman L, Hansen M. The use of evidence-based practice by occupational therapists. *Occup Ther Health Care*. 2009;23(2):134-145.

30. Stronge M, Cahill M. Self-reported Knowledge, Attitudes and Behaviour towards Evidence-based Practice of Occupational Therapy Students in Ireland: Evidence-based Practice, Irish OT students. *Occup Ther Int*. 2012;19(1):7-16. doi:10.1002/oti.328

31. McEvoy MP, Williams MT, Olds TS. Development and psychometric testing of a trans-professional evidence-based practice profile questionnaire. *Med Teach*. 2010;32(9):e373-e380.

32. McEvoy MP, Williams MT, Olds TS, Lewis LK, Petkov J. Evidence-based practice profiles of physiotherapists transitioning into the workforce: a study of two cohorts. *BMC Med Educ*. 2011;11(1):100.

33. Green C, Brazier J, Deverill M. Valuing health-related quality of life. *Pharmacoeconomics*. 2000;17(2):151-165.

34. Weernink MGM, Janus SIM, van Til JA, Raisch DW, van Manen JG, IJzerman MJ. A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making. *Pharm Med*. 2014;28(4):175-185. doi:10.1007/s40290-014-0059-1

35. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *PharmacoEconomics*. 2017;35(1):21-31. doi:10.1007/s40273-017-0545-x

36. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846-855. doi:10.1016/j.jval.2015.07.004

37. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Econ*. Published online 1999:11.

38. Brazier J, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32). doi:10.3310/hta16320

39. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. *Qual Life Res*. 2010;19(6):907-917. doi:10.1007/s11136-010-9646-0

40. Poissant L, Mayo NE, Wood-Dauphinee S, Clarke AE. The development and preliminary validation of a Preference-Based Stroke Index (PBSI). *Health Qual Life Outcomes*. Published online 2003:15.

41. Kuspinar A. *Development of a Preference-Based Measure for Multiple Sclerosis: The Preference-Based Multiple Sclerosis Index (PBMSI)*. McGill University; 2014. Accessed May 3, 2019. https://onlinelibrary.wiley.com/doi/abs/10.1111/scs.12482

42. Moga AM, Twells LK, Mayo NE. The prototype of a preference-based index of weight-related quality of life: demonstrating the possibilities. *Qual Life Res*. Published online May 24, 2022. doi:10.1007/s11136-022-03156-5

43. Green PE, Rao VR. Conjoint measurement-for quantifying judgmental data. *J Mark Res*. 1971;8(3):355-363.

44. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403-413.

45. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol*. 2015;3(1):1-12.

46. Powell BJ, Beidas RS, Lewis CC, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. *J Behav Health Serv Res*. 2017;44(2):177-194. doi:10.1007/s11414-015-9475-6

47. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*. Cambridge university press; 2000.

48. Viney R, Lancsar E, Louviere J. Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Rev Pharmacoecon Outcomes Res*. 2002;2(4):319-326.

49. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *PharmacoEconomics*. 2008;26(8):661-677. doi:10.2165/00019053-200826080-00004

50. Johnson RM. Trade-off analysis of consumer values. *J Mark Res*. 1974;11(2):121-127.

51. Waltz TJ, Powell BJ, Chinman MJ, et al. Expert recommendations for implementing change (ERIC): protocol for a mixed methods study. *Implement Sci*. 2014;9(1):1-12.

52. Cunningham CE, Barwick M, Short K, et al. Modeling the mental health practice change preferences of educators: A discrete-choice conjoint experiment. *School Ment Health*. 2014;6(1):1-14.

53. Williams NJ, Candon M, Stewart RE, et al. Community stakeholder preferences for evidence-based practice implementation strategies in behavioral health: a best-worst scaling choice experiment. *BMC Psychiatry*. 2021;21(1):74. doi:10.1186/s12888-021-03072-x

54. Goossens A, Bossuyt PM, de Haan RJ. Physicians and nurses focus on different aspects of guidelines when deciding whether to adopt them: an application of conjoint analysis. *Med Decis Making*. 2008;28(1):138-145.

55. Whitty JA, Oliveira Gonçalves AS. A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best–Worst Scaling for Eliciting Preferences in Healthcare. *Patient - Patient-Centered Outcomes Res*. 2018;11(3):301-317. doi:10.1007/s40271-017-0288-y

56. Cheung KL, Wijnen BFM, Hollin IL, et al. Using Best–Worst Scaling to Investigate Preferences in Health Care. *PharmacoEconomics*. 2016;34(12):1195-1209. doi:10.1007/s40273-016-0429-5

57. Orme BK. Getting started with conjoint analysis: strategies for product design and pricing research. Published online 2006.

58. Marley AAJ, Louviere JJ. Some probabilistic models of best, worst, and best–worst choices. *J Math Psychol*. 2005;49(6):464-480. doi:10.1016/j.jmp.2005.05.003

59. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-189. doi:10.1016/j.jhealeco.2006.04.002

60. Lancsar E, Louviere J, Donaldson C, Currie G, Burgess L. Best worst discrete choice experiments in health: Methods and an application. *Soc Sci Med*. 2013;76:74-82. doi:10.1016/j.socscimed.2012.10.007

61. Krucien N, Watson V, Ryan M. Is Best–Worst Scaling Suitable for Health State Valuation? A Comparison with Discrete Choice Experiments. *Health Econ*. 2017;26(12):e1-e16. doi:10.1002/hec.3459

62. Johnson RM, Orme BK. A new approach to adaptive CBC. In: *Sawtooth Software Conference*. Citeseer; 2007:85.

63. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health*. 2013;16(1):3-13. doi:10.1016/j.jval.2012.08.2223

64. Johnson R. Understanding HB: An Intuitive Approach. Sawtooth Software. Published 2000. Accessed April 26, 2022. https://sawtoothsoftware.com/resources/technical-papers/understanding-hb-an-intuitive-approach

65. Sawtooth. The CBC/HB System Technical Paper V5.6. Published 2009. Accessed April 26, 2022. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-technical-paper

66. Howell J. CBC/HB for Beginners. Sawtooth Software. Published 2009. Accessed April 20, 2022. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-for-beginners

67. Lizarondo LM, Grimmer-Somers K, Kumar S, Crockett A. Does journal club membership improve research evidence uptake in different allied health disciplines: a pre-post study. *BMC Res Notes*. 2012;5(1):588. doi:10.1186/1756-0500-5-588

68. Wilkinson SA, Hinchliffe F, Hough J, Chang A. Baseline Evidence-based Practice Use, Knowledge, and Attitudes of Allied Health Professionals: A Survey to Inform Staff Training and Organisational Change. *J Allied Health*. 2012;41(4):177-184.

69. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, USA; 2015.

70. Smits PA, Champagne F, Brodeur JM. A mixed method study of propensity for participatory evaluation. *Eval Program Plann*. 2011;34(3):217-227. doi:10.1016/j.evalprogplan.2011.03.002

71. Bennebroek Gravenhorst KM, Werkman RA, Boonstra JJ. The change capacity of organisations: general assessment and five configurations. *Appl Psychol*. 2003;52(1):83-105.

72. Myers CT, Lotz J. Practitioner Training for Use of Evidence-Based Practice in Occupational Therapy. *Occup Ther Health Care*. 2017;31(3):214-237. doi:10.1080/07380577.2017.1333183

73. Rochette A, Brousseau M, Vachon B, Engels C, Amari F, Thomas A. What occupational therapists' say about their competencies' enactment, maintenance and development in practice? A two-phase mixed methods study. *BMC Med Educ*. 2020;20. doi:10.1186/s12909-020-02087-4

74. Condon C, McGrane N, Mockler D, Stokes E. Ability of physiotherapists to undertake evidence-based practice steps: a scoping review. *Physiotherapy*. 2016;102(1):10-19.

75. Robertson L, Graham F, Anderson J. What Actually Informs Practice: Occupational Therapists' Views of Evidence. *Br J Occup Ther*. 2014;76:317-324. doi:10.4276/030802213X13729279114979

76. Jeffery H, Robertson L, Reay K. Sources of evidence for professional decision-making in novice occupational therapy practitioners: clinicians' perspectives. *Br J Occup Ther*. 2020;84:030802262094139. doi:10.1177/0308022620941390

77. Robertson L, Graham F, Anderson J. What Actually Informs Practice: Occupational Therapists' Views of Evidence. *Br J Occup Ther*. 2013;76(7):317-324. doi:10.4276/030802213X13729279114979

78. Lyons C, Brown T, Tseng MH, Casey J, McDonald R. Evidence-based practice and research utilisation: Perceived research knowledge, attitudes, practices and barriers among Australian paediatric occupational therapists. *Aust Occup Ther J*. 2011;58(3):178-186.

79. McCluskey A, Vratsistas-Curto A, Schurr K. Barriers and enablers to implementing multiple stroke guideline recommendations: a qualitative study. *BMC Health Serv Res*. 2013;13(1):323. doi:10.1186/1472-6963-13-323

80. Devlin NJ, Hansen P, Selai C. Understanding health state valuations: a qualitative analysis of respondents’ comments. *Qual Life Res*. 2004;13(7):1265-1277.

81. Lezak MD. Executive functions and motor performance. *Neuropsychol Assess*. Published online 1995:650-685.

82. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. Published online 2003:582-592.

83. Sočan G. Assessment of reliability when test items are not essentially τ-equivalent. *Dev Surv Methodol*. 2000;15:23-35.

84. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas*. 2004;64(3):391-418.

85. Glasgow RE. What Does It Mean to Be Pragmatic? Pragmatic Methods, Measures, and Models to Facilitate Research Translation. *Health Educ Behav*. 2013;40(3):257-265. doi:10.1177/1090198113486805

86. Swait J, Adamowicz WL. *The Effect of Choice Environment and Task Demands on Consumer Behavior: Discriminating between Contribution and Confusion*.; 1996.

87. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev*. 1956;63(2):81-97. doi:10.1037/h0043158

88. Payne JW, Payne JW, Bettman JR, Johnson EJ. *The Adaptive Decision Maker*. Cambridge university press; 1993.

89. Wankah P, Guillette M, Dumas S, et al. Reorganising health and social care in Québec: a journey towards integrating care through mergers. *Lond J Prim Care*. 2018;10(3):48-53. doi:10.1080/17571472.2018.1453957

Appendices

Appendix I. PIRE-CDMI inter-item polychoric correlation matrix

	Use of research evidence	Self-efficacy	Resources	Effort	Keep up to date
Use of research evidence	1				
Self-efficacy	0.53	1			
Resources	0.5	0.17	1		
Effort	0.02	0.21	0.68	1	
Keep up to date	0.64	0.49	0.48	0.17	1

Chapter 10: Discussion

This chapter begins with a summary of the four studies of this doctoral research in relation to the objective of this thesis. The summary is followed by a discussion of the contributions of my doctoral research on a theoretical, methodological, and practical level. Finally, I highlight the strengths and limitations of this doctoral research, propose areas for future research and provide a concluding statement.

Summary of findings

There is an ongoing need for effective rehabilitation services[1]. Evidence-based practice (EBP) in rehabilitation can optimize quality of care through the creation and application of research evidence on the effectiveness of rehabilitation approaches and can contribute to reducing ineffective and potentially harmful practices[1–3]. EBP is defined as a complex and multidimensional approach to clinical decision-making (CDM) that combines the best available research evidence, clinical expertise, and patients' preferences and values, within the context of available resources[4–6]. EBP is not only a core professional competency for occupational (OTs) and physical (PTs) therapists but an expectation from consumers and funders who demand and deserve quality services based on best available evidence[7].

Given the call for EBP, robust measurement practices are essential to identify whether clinicians are exercising EBP, which factors related to EBP could be improved and what new or existing practices need to be implemented or discontinued[8,9]. Information generated from EBP measures can demonstrate the effectiveness of EBP interventions, highlight areas for future research, and allow for a targeted allocation of resources to better equip clinicians in embracing and adopting EBP.

While more than 30 measures have been developed and used to measure EBP with OTs and PTs[9,10] in the past two decades, the measurement of EBP is replete with challenges: (1) there is a shortage of comprehensive questionnaires representing the most salient individual *and* organizational EBP domains[9]; (2) ordinal data are treated as interval-level data[11–13]; (3) the interpretation of EBP is fragmented due to "profile-type" measures which summarize results of domains independently[14]; and (4) the relative importance of the domains that influence EBP and how these can be quantitatively captured in a measure are unknown. The research presented in this doctoral dissertation addresses these gaps.

The overall objective of this doctoral research was to develop a brief, multidimensional index of a rehabilitation clinician's propensity to integrate research evidence into CDM, also called the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). For the purposes of this research, propensity to integrate research evidence is defined as the probability of a rehabilitation clinician to integrate research evidence into CDM given individual and organizational determinants. A propensity score is the conditional probability of an individual having an exposure given a multitude of variables[14,15].

To operationalize the overarching aim of my doctoral work, I conducted four research projects from which I have prepared five manuscripts. The aim of the first study was to identify and describe the methods, results, recommendations, reported challenges and areas for future practice across systematic reviews (SRs) on EBP measures in healthcare. This umbrella review included 10 SRs. The review included recommendations for the development of future measures and served to inform the next phase of my research, the development of the index. The results from this study were reported in two manuscripts (*Manuscripts 1 and 2)*.

Manuscript 1, which was published in *JBI Evidence Synthesis* in 2022, identified 204 existing EBP measures across the healthcare professions and described 27 measures deemed adequate in the SRs. In rehabilitation, only one of three SRs explicitly recommended eight measures for use. None of these eight recommended measures comprehensively span all EBP domains of knowledge, skills, attitudes, and behaviors. More so, none include items related to the organization and available resources.

Manuscript 2, which reported on the challenges, recommendations for practice and areas of future research in the measurement of EBP, was published in the *Journal of Evaluation in Clinical Practice* in 2022. The paper identified important limitations with existing EBP measures including the limited psychometric testing as it pertains to validity, reliability and feasibility/acceptability, the lack of construct clarity, and an overrepresentation of certain steps of the EBP process (i.e., appraise) at the expense of other ones (i.e., apply and assess). In response to the reported challenges, authors of SRs recommend that researchers (1) acknowledge the multidimensionality of EBP by developing comprehensive measures and (2) place more emphasis on the feasibility and acceptability of future measures. In sum, this study suggests that existing measures may be insufficient in capturing the multidimensional, contextual, and dynamic nature of EBP substantiating the need for a robust, comprehensive, and clinically relevant measure of EBP.

Manuscript 3 reported the results of study 2 which aimed to describe the initial steps required to develop the *Propensity to Integrate Research Evidence into Clinical Decision-Making Index* (PIRE-CDMI). This study consisted of a secondary analysis of an existing dataset that aimed to measure and understand how EBP evolves from graduation and into practice over a period of three years in a sample of 257 newly graduated OTs and PTs from 29 rehabilitation programs in Canada[16,17]. In this previous study by Al Zoubi et al., six EBP domains (knowledge, use of research evidence, self-efficacy, resources, attitudes, and activities related to EBP) were identified through a literature search and nominal group technique[16]. Using the results from Rasch Analysis on the data of 127 Canadian OTs and PTs who completed a total of 70 items from six unidimensional measures of EBP, one item was selected per EBP domain to be included in the prototype index. Best performing items were selected by visual inspection of the threshold maps (based on the work of Brazier et al.[18] and Young et al.[19]) and expert consensus. The prototype index comprised five items, with three response levels per item. Testing of the prototype index demonstrated that it behaves consistently across age groups, gender, and settings, and provides comparable information to other unidimensional EBP measures in a succinct way[20]. This study provided a proof-of-concept measure and evidence that such a measure would be valuable.

Manuscript 4 reported on a three-phased study aimed at revising and rewriting the five items included in the prototype index in English and French. In the first phase, I conducted an online focus group with rehabilitation clinicians and experts in the field of EBP (both likely end users of the index) and cognitive interviews. Based on participant feedback, important modifications were made for clarity and interpretability of items and response options. The second phase consisted of 24 online cognitive interviews with OTs and PTs whereby end-users of the index were asked to elaborate on their interpretation on the meaning of the items, the comprehensibility of items, and the appropriateness of response options. The interviews led to 12 iterations of changes to the index to increase the clarity and interpretability of items and reduce measurement error. The last phase consisted of an online cross-sectional survey with English and French-speaking healthcare students and professionals to demonstrate the equivalency of response options in both languages. The survey required respondents to interpret common words (response option labels) and rate them on a numerical scale from 0 to 100. As such, respondents were not excluded based on profession or level of training. Regression analysis suggested no important difference between response options in English and French across items.

Manuscript 5 described the process used to elicit weights for each dimension and dimension-level in the index using a best-worst scaling choice exercise. Using a software called *Sawtooth*[21], OTs, PTs, clinical managers, and experts in EBP were asked to select "the best" and "the worst" among three hypothetical profiles containing varied levels of the five EBP dimensions. The choices made by respondents were reflective of the relative importance of certain EBP domains over others. Based on the results from the best-worst scaling exercise, part-worth utilities were estimated for each response level for all five items and a scoring algorithm was developed for the index which allowed for a mathematically and theoretically valid total score to be generated. Index scores span a continuum from 0 (neutral propensity to integrate research evidence) to 100 (high propensity for EBP). When interpreting the results from the index, a score of 0 designates neutral propensity rather than suggesting no propensity of integrating research evidence. Propensity is built with each item level-increment. This way, the index leverages a respondent's strengths rather than their weaknesses. A more positive framing may help clinicians perceive the index as a constructive tool for their professional development rather than an ominous one. This study also contributes validity evidence regarding internal structure, relationships to other variables and feasibility which support the use of the PIRE-CDMI to draw inferences about OTs and PTs in Canada.

Theoretical contributions

The research presented in this dissertation highlights limitations with existing approaches to measuring EBP and contributes to redefining how EBP has traditionally been measured by proposing a novel approach using the concept of *propensity*. I make the case for using propensity variables related to the integration of research evidence into CDM based on two main arguments. First, using *propensity*, the conditional probability for a behavior[14], acknowledges the dynamic and multidimensional nature of the EBP process which varies with each clinical encounter. This responds to a limitation in current measures which assume that EBP indicators have static properties. Second, the PIRE-CDMI focuses on propensity to integrate research evidence into

CDM (an important element of EBP) because there exists a well-articulated, accepted body of knowledge with respect to the influence of certain individual and organizational factors on the integration of research evidence into CDM and to the measurement of those factors[22]. The propensity variables used are well-established in the literature and can be characterized through set features[22].

Most of the existing literature claiming to measure EBP does not account for sources of knowledge other than research evidence. Unless these other sources of knowledge are accurately captured, it is a conceptual flaw to call these *EBP* questionnaires or measures. Given the immensurability of non-propositional forms of knowledge such as tacit knowledge[23,24], EBP cannot be measured directly, but can only be estimated and inferred with related variables[25].

The above-mentioned justifications for using *propensity* are closely related to the philosophical understanding of EBP from a social constructivist perspective whereby clinicians draw on multiple sources of knowledge in socially situated contexts to form clinical decisions. Measuring EBP is an epistemic activity in two ways: first, the integration of different sources of knowledge to make decisions involves distinguishing what constitutes a justifiable belief; second, measurement involves judging the best way to gain knowledge about a phenomenon. Despite the resounding presence of epistemic assumptions in this area of research, the literature reporting on EBP measures fails to discuss the philosophical tensions inherent in measuring EBP. It is anticipated that this dissertation can contribute to future discussions regarding the operationalization of EBP. In identifying and describing these epistemic tensions, I hope to honor the methodological commitment, "the imperative of progress"[26(p44)], which suggests that scientists continue to improve the epistemic virtues of their predecessors. This research highlights an area that requires increased attention so that the field of measuring EBP can advance with philosophical alignment.

The findings presented in *Manuscript 5* reveal, for the first time in rehabilitation, an estimate of the relative importance of five EBP dimensions on a clinician's propensity to integrate research evidence into practice. These new results advance knowledge on the factors that influence EBP by quantifying the relationships between these factors. Notably, two dimensions, *use of research evidence* and *keep up to date*, had higher relative importance compared to the other three dimensions, *self-efficacy*, *effort*, and *resources*. These findings on the relative importance of EBP dimensions have implications for future studies aiming to optimise EBP. Researchers and stakeholders involved in the creation and application of knowledge could prioritize evidence-based activities such as keeping up with research evidence over other factors such as self-efficacy which are rated lower in average importance.

Further, this dissertation has considered the emerging discourse of validity as a social imperative, a perspective which has not been previously discussed in relation to the measurement of EBP. Applying the concept of validity as a social imperative entails using a holistic lens to examine the individual and social consequences of measurement[27]. This consideration is similar to what Messick describes as *consequential validation*, where the consequences of the decisions based on a test must be factored into the evaluation of validity[28,29]. EBP measurement practices are a social act which includes the assessor (or researcher) and the clinician, and are value-laden, in that a message is communicated to respondents about what matters and the types of professional behaviors that are valued[30]. Given that EBP is a professional responsibility, the measurement of EBP indicators can introduce an intimidating power dynamic wherein the clinician fears disciplinary consequences that may jeopardize their professional credibility (and potentially their right to practice). Considering these potentially negative consequences, clinicians may resort to distorting their responses to ones that are perceived to be more socially acceptable. Throughout this dissertation, the measurement of EBP is positioned as an approach to facilitate professional development rather than a test of professional competence. In this sense, the PIRE-CDMI becomes a support *for* clinicians rather than an evaluation *of* clinicians. This discussion advances the field of EBP measurement by highlighting the responsibility of measure developers to examine the social consequences of testing to obtain more valid inferences, increase the social utility of measures and support ethical evaluation practices.

Methodological contributions

The umbrella review (study 1) was the first of its kind to comprehensively review the literature on EBP measures across the healthcare professions and acts as one of few umbrella reviews synthesizing psychometric SRs[31,32]. Findings from *Manuscript 1* highlight a need to improve the quality of systematic reviews on EBP measures regarding the search strategy, quality assessment of primary studies and of measures, and adherence to SR methodological guidelines and reporting standards. The work contained within *Manuscript 1* also reveals that different methodological approaches were used by authors of SRs to assess the adequacy of EBP measures, highlighting heterogeneous practices. Most authors (except one group[33]) followed diverse standards for patient-reported outcome measures wherein construct, content and criterion validity are considered discrete entities. As described in chapter 4, this dissertation embraces a different approach to validity and adheres to the contemporary perspective of validity as proposed by *The Standards of Educational and Psychological Testing*[20] rather than the trinitarian point of view often referenced in the clinical sciences[34]. The unified conceptualization of validity, whereby all types of validity are considered to be construct validity, was first proposed by Messick in the 1970s[28,35]. Although this approach is labelled *contemporary*, this conceptualization of validity was put forward almost 50 years ago, and disciplines outside of psychology and education have had major difficulties claiming these concepts[34,36,37]. In this view, validity is regarded as a property of scores and score interpretations rather than a property of the measure itself[20,35,38]. The issue with referring to the "validity and reliability *of measures*", as was the case in eight of the 10 included SRs in *Manuscript 1*, is that validity is perceived as a property of a measure rather than a property of the inferences that can be made about a set of people in a given context. The major concern with understanding validity as a property of the measure is the dichotomized conclusions that a measure can be valid (or invalid) which assumes that a measure receives a "gold seal of approval"[39(p858)]. This approach provides a "false sense of security for assessment practitioners"[39(p859)] and may hinder them from conducting an appraisal of the existing validity evidence as it applies to their context. It is hoped that the ideas presented in *Manuscript 1* regarding the application of validity theory to the measurement of EBP can (1) serve as a basis for learning, reflection, and discussion on the different validity discourses and (2) highlight the importance of being explicit about one's specific position regarding validity.

Results from *Manuscript 2* suggest a need for diversification of methodological approaches including qualitative methodologies to obtain a richer understanding of EBP. Thus, the self-report index developed in the context of this dissertation can serve as one method of drawing inferences related to EBP which can be complemented with other methodologies such as interviews or observation. This finding suggests expanding actual EBP measurement practices, which most often consist of a single approach (e.g., a self-report questionnaire), by adding other sources of information for data triangulation. Doing so would help corroborate findings between measurement approaches and reveal any weaknesses in methods of data collection, improvements that would advance the field of EBP measurement. Further, diversifying the measurement approach could help ease the epistemological tensions in measuring such a complex process.

The research presented in this dissertation makes a key methodological contribution by using novel and innovative approaches to develop the first multidimensional index in EBP in rehabilitation, but also across the healthcare professions. Of note, modern measurement theory, specifically Rasch Measurement Theory (RMT), was employed to ensure that the items selected in the index had interval-level properties allowing for a mathematically valid summing of items[40,41]. This responds to a serious concern identified regarding EBP measures whereby arithmetic operations and parametric inferential statistics are conducted on ordinal data[11–13]. RMT has only been used on three occasions in nursing and once in rehabilitation to develop EBP measures. In nursing, RMT has been used to develop measures of capability beliefs on EBP for students and practitioners[42], self-efficacy to enact EBP in practice for graduating students[43], and EBP knowledge for use with undergraduate and graduate students[44]. In rehabilitation, Al Zoubi and colleagues[16] developed the 55-item multidimensional measure wherein mathematically valid total scores could be generated for each of the six EBP domains. The research presented in *Manuscript 3* builds on and expands the work by Al Zoubi et al. and demonstrates how RMT and expert consensus can be leveraged to reduce 55 items to 5 items.

To ensure that the five items selected for inclusion in the index were pertinent, clear, and retained their interval-level properties, I carried out a comprehensive process of revising and rewriting the items with end-users, completely online (due to the COVID-19 pandemic). In the study reported in *Manuscript 4*, it was paramount to ensure that the items were best representative of the actual process of integrating research evidence into CDM that occurs in clinical practice. The findings from 24 cognitive interview led to three major item modifications, among minor changes: first, participants consistently reported that the item measuring attitudes towards EBP was prone to social desirability bias despite multiple modifications. To reduce measurement bias, this item was modified to focus on the perceived effort required to integrate research evidence into practice as a proxy for attitudes or perceived value of EBP. Social psychologists and behavioral theorists have identified effort as largely contributing to behavioral motivation[45–47]. The second important item modification consisted of introducing the idea of a "clinical uncertainty" to contextualize the behavior of seeking and integrating research evidence into CDM. Adding this trigger was essential for eliciting the behavior of interest whilst recognizing that CDM has an element of automaticity: over time, research-based knowledge may no longer be distinguished as *research evidence* if it becomes consolidated into, and/or combined with tacit or experiential knowledge. Third, participants voiced how they gathered research evidence from various sources outside of scientific journals. For this reason, it was essential that the index be inclusive of various sources of research evidence such as consultation with peers and email notifications.

Given Canada's linguistic diversity, methods to ensure the equivalence of questionnaire versions in different languages are warranted to decrease systematic differences between language groups. The simultaneous translation approach was employed during the item rewriting process and has been recognized for its efficiency and effectiveness in translating questionnaires in comparison to the traditional forward-backward translation method[48–50]. Despite best efforts, translation into other languages is complexified by linguistic and cultural differences such as idiomatic expressions. To exercise due diligence following the rewriting of items, I conducted a cross-sectional survey to quantitatively compare the English and French response options on a 0-100 scale and to demonstrate that respondents are interpreting items in a similar fashion. This validation method is a useful approach to comparing textual labels in different languages and this study serves as one of few examples in the literature[51,52].

This doctoral research advances the measurement of EBP by allowing for a single global score to be generated that captures multiple domains and considers the gains in one domain against losses in another. I adopted the methodology associated with the development of preference-based measures for patients[53–55] and applied it to an entirely new context: EBP with rehabilitation clinicians. Reported in *Manuscript 5,* this approach of developing a weighted, multidimensional index to infer EBP has never been proposed in rehabilitation, nor across the health professions. Doing so addresses a methodological gap in current measures wherein the interdependency between factors known to influence EBP is discounted, and results are summarized separately by domain.

Specifically, this research used choice-based experiments (best-worst scaling) to elicit weights for each dimension. In doing so, this study provides methodological insights on the feasibility and acceptability of an online weight elicitation method from the perspectives of rehabilitation clinicians, clinical managers, and experts in EBP. More importantly, the use of choice-based methods allows for the first estimate of the relative importance of five EBP domains from stakeholders' perspectives. The field of health services research and knowledge translation (KT) could benefit from using choice experiments such as best-worst scaling to derive preferences from groups of individuals and this study provides a detailed example.

The PIRE-CDMI was developed to provide researchers with a robust clinical outcome measure to identify EBP engagement levels in practice, create interventions and assess their effectiveness, or measure changes in EBP over time. The development of the PIRE-CDMI responds to a call for multidimensional and pragmatic measures to support EBP research in rehabilitation[8,56]. Pragmatic measures have been characterized as being low in burden for respondents and administrators, broadly applicable, actionable, and unlikely to cause harm[56].

The PIRE-CDMI overcomes one of the major shortcomings with existing EBP measures: the failure to combine EBP domains for an overall multidimensional indicator of EBP. The PIRE-CDM allows to identify specific areas for improvement by way of individual dimension results, but also provides a total score allowing for a broad interpretation of one's propensity to integrate research evidence into CDM.

The propensity score can be used as a weighting adjustment procedure for matching and stratification to account for group differences and provide more precise estimates when estimating the effects of interventions in EBP studies[15,57]. Specifically, the index can be added to a baseline questionnaire on clinician demographics. Then, clinicians' propensity scores could be (1) incorporated into study design through matched sampling whereby clinicians with similar propensity scores could be separated into a control and intervention group (or different intervention groups); (2) incorporated into purposive sampling procedures to gain a homogenous or heterogenous sample based on propensity scores; or (3) integrated into modelling the effect of an intervention (or intervention parameters) through stratification by sub-groups based on propensity scores.

The research reported in this dissertation also has implications for KT studies. The thriving field of KT, which addresses a need to reduce research to practice gaps, upholds similar principles to EBP relating to the use of empirical research to improve health services[58,59]. The PIRE-CDMI can be used as an outcome measure in KT effectiveness studies of interventions focused on the implementation of EBPs with rehabilitation professionals. Assessing clinician characteristics and organisational culture prior to designing interventions is crucial as illustrated in frameworks such as the *Promoting Action on Research Implementation in Health Services (PARIHS)* framework[60], *The Consolidated Framework for Implementation Research (CFIR)*[61] and the *Exploration, Preparation, Implementation, Sustainment (EPIS)* framework[62]. For example, in the *EPIS*[62], the exploration phase entails assessing provider (e.g., clinician) characteristics such as education level, experience with EBP, dispositional innovativeness and attitudes towards EBP. The PIRE-CDMI can be used in this capacity.

Further testing is necessary to substantiate the use of the index as an outcome measure in clinical research (details are presented in the subsection entitled *Future directions*).

Practice contributions

In addition to its potential use as an outcome measure in research, the PIRE-CDMI can be used in clinical practice as a catalyst for self-reflection and quality improvement. Although more testing is needed to ensure that the index can be used as a self-reflection tool (see the section *Future Directions* for more information), the index is likely to be equally useful for clinical managers who may use results periodically with clinicians to initiate discussion about professional development, areas for improvement and necessary organizational supports that should be put in place to contribute to better service provision.

The index was not developed to replace comprehensive unidimensional measures, but to provide a feasible and acceptable alternative that takes busy clinicians about one minute to complete (as demonstrated in chapter 9). This index can be rapidly deployed as a global screen of a clinician's propensity to integrate research evidence which can then be complemented, if necessary, with more comprehensive and lengthier measures. The index can signal areas for improvement pertaining to the five included dimensions (use of research evidence, self-efficacy, resources, effort, and keep up to date) impacting the integration of research evidence into CDM. Strategies can then be developed to target the specific areas requiring support. However, the index was not developed to be used for regulatory or disciplinary purposes. If used in such high-stakes contexts, clinicians may feel obliged to respond in a manner that avoids social consequences; this will considerably increase the risk for social-desirability bias and inaccurate findings[63,64]. To avoid feelings of coercion and an increased risk of measurement bias, it is essential to clearly state that this index is not designed for auditing of professional practice[65]. Although it is not recommended to use this index in high-stakes contexts, the precise scores generated can support professional growth and reflection by providing clinicians with a numerical indicator of their current propensity.

While future testing will contribute to the validity argument in relation to other variables, it is not expected that there will be a meaningful cut-off score. The primary aim of this index is not to classify clinicians into groups. In fact, the interpretation of propensity scores exists on a spectrum. Further, a dichotomized approach to interpreting results from this index, such as claiming "good" versus "bad" propensity to integrate research evidence, assumes that clinical performance is easily divided into those who can and those who cannot rather than reflecting a multifaceted process of CDM that changes with each encounter. This aligns with Kleiner et al. who conducted an integrative review of the qualities of a good physiotherapist and discussed "a need to overcome dichotomous thinking, and to re-conceptualize assumed dualities as being intertwined"[66(p18)]. Considering propensity on a spectrum (from neutral to high) can help rehabilitation clinicians reconceptualize EBP as one important, on-going element in the wider spectrum of professional practice, rather than a binary act of doing EBP or not doing EBP.

Strengths and limitations

Strengths

Methodological rigor

First, to inform the measure development process and substantiate the need for a new measure, an umbrella review was conducted following JBI methodological guidelines. This review was published in *JBI Evidence Synthesis* and was promoted as an exemplary umbrella review by the editor in chief, Dr. Eduoardo Aromataris, on Twitter[67]. Some notable characteristics include the development of a comprehensive search strategy in collaboration with a health services librarian, the quality appraisal process of included reviews and the extraction of data by two independent reviewers. Second, the development of the index is grounded in modern measurement theory (Rasch Analysis) which is increasingly recognized for its utility as an essential and robust approach to analysis of categorical scales[19,68–72]. Third, the elicitation of part-worth utilities by best-worst scaling was conducted using state-of-the-art software (Sawtooth Software) aligned with current best practices in the field[73,74]. Finally, this dissertation conceptualized validity as a unified concept based on the *Standards* which are considered the gold standard for assessment practices in the social sciences (e.g., education, psychology, and employment)[20]. In addition to the validity evidence based on content, internal structure and relations to other variables is provided throughout *Manuscripts 3 to 5*, important efforts were made in this dissertation to describe the feasibility and acceptability of the index and the consequences of testing, considerations that are often overlooked in the literature but are deemed important by the *Standards*[20].

Development of an index with and for end-users

In considering validity as a social imperative, the PIRE-CDMI was designed as an evaluation *for* clinicians rather than *of* clinicians. To minimize social desirability bias and negative consequences for clinicians, it was essential to (1) clarify that the purpose of the index is not to evaluate clinicians' competency for regulatory action but to administer a quick screen which can help identify and support their needs, and (2) develop the index alongside a range of potential end-users of the index to ensure content relevance and minimize measurement bias and respondent burden.

The interpretation of results from the index builds on the strengths rather than weaknesses as respondents score on a continuum from neutral (Score of 0/100) to high propensity (Score of 100/100). As such, the focus is shifted from what doesn't work (e.g., barriers to EBP) to what works and can be improved. Framing results in this way can be perceived as an empowering approach to measurement[75] and reduces the chance that using this index will cause harm to clinicians (i.e., produce unintended consequences)[56].

Increasingly both in implementation science and in measurement, best practices warrant the participation of end-users in the research process to enhance the relevance and appropriateness of research[49,76,77]. In the context of this dissertation, I deployed four discrete recruitment endeavors to reach a broad sample of: (1) Canadian OTs and PTs and content experts in EBP for the focus group; (2) Canadian OTs and PTs for cognitive interviews; (3) Canadian English and French-speaking healthcare students and professionals for the translation validation survey; and (4) Canadian OTs, PTs, clinical managers, and experts in EBP for the best-worst scaling exercise. The online communication tools enabled the inclusion of participants across Canada which increases the generalizability of results. Throughout this research, participants offered indispensable and insightful feedback which helped to shape the propensity index into a more meaningful and relevant measure for clinicians. The participation of clinicians in providing feedback on the interpretability of the index strengthens the validity argument that this index measures what it purports to measure. Furthermore, an important overall strength of this study was the positionality of myself and the research team as OT/PT researchers in the fields of EBP and KT. The development of the PIRE-CDMI builds on the expertise of clinicians, clinical managers, experts in EBP and the research team.

Limitations

In addition to the specific limitations described in the individual manuscripts, two main limitations apply to this dissertation on a broader level.

First, there exists a delicate trade-off between brief practical scales and lengthier, more comprehensive scales. Short scales have been criticized for not adequately covering the construct of interest and compromising the validity of inferences[78,79]. To respond to a call for pragmatic measures that are brief, broadly applicable and have a low burden on administrators and respondents[56] and in light of the resource-strained healthcare context, this research focused on brevity to maximize usability and increase chances for high response rates[80,81]. Given the rigorous measure development process reported in this dissertation, we do not believe to have compromised the methodological quality for usability[82]. This index serves a specific purpose and does not aim to replace lengthier measures. In the future, one could aim to resolve this tension by developing a longer version of this index.

Second, the dimensions included in this index are based on previous work which has identified the core EBP domains[16]. Although other domains not included in this index may also contribute to the construct of interest such as the attitudes of colleagues on the importance of evidence[83], the identification of the domains included in this index is based on a through literature review and nominal group process conducted by prominent scholars in the field of measurement and EBP. Future work can expand on this study to provide additional insights on domains contributing to a clinician’s propensity to integrate research evidence into CDM. An important consideration for this future work is to ideally keep the item count lower than 10 for the measure to remain practical[8]. Further, the number of dimensions and dimension-levels must be kept at a minimum for choice-based weight elicitation exercises; it is recommended to limit the number of domains in model simulations to the six most important[84].

Future directions

The research presented in this dissertation has offered avenues for future research. In addition to future directions described in *Manuscript 1* (i.e., an empirical analysis of the theoretical tensions regarding the conceptualization of validity in EBP; developing a practical resource to select an EBP measure for a specific use; extending the methodological guidance for the conduct of umbrella reviews specific to psychometric SRs). I will discuss two broad research avenues.

The first broad research avenue entails further field testing of the index as an outcome measure with different samples of OTs and PTs, before inferences can be drawn with confidence. Future studies could estimate the (1) relationship between index scores and comprehensive measures of EBP; (2) relationship between index scores and other indicators relating to the integration of research evidence into CDM (e.g., frequency of consulting scientific journals); (3) sensitivity to change after an EBP intervention; and (4) associations with variables such as time since graduation, profession, clinical setting and area of practice, workload, and peer isolation. As stated in *The Standards*, further testing and adaptations can also consider the accessibility of the index to ensure accommodation and fairness in testing[20]. Future research could also aim to confirm the dimensional structure of the PIRE-CDM through confirmatory factor analysis[85] and to verify that the interval properties of the scale still hold through Rasch analysis. Additionally, the additive model generated using the best-worst scaling technique and hierarchical Bayes statistical inference in *Manuscript 5* could be empirically compared to a multi-attribute utility function. A different approach to estimating weights could be used such as direct elicitation of numerical ratings with a visual analogue scale[86] (1) for each dimension separately and (2) for corner states - a corner state is when one dimension is set at its worst level while all other dimensions are set at best.

A second broad research avenue consists of expanding the use of this index into a self-reflection tool for OTs and PTs. Although rehabilitation regulatory bodies and national organizations set high-level expectations for EBP, the uptake of research evidence into CDM is lagging[17,87–95]. This may be related to the lack of specific and regular feedback regarding clinicians' enactment of the EBP process. The index has the potential to be used as a professional self-reflection tool to provide clinicians with an indicator of their engagement with research evidence and allow them to be more actively engaged in their role as life-long learners and evidence-based clinicians. Self-monitoring of behavior is a well-known strategy for behavior change[96,97]. Like a formative assessment in education, a tool designed to foster reflection on one's clinical practice can be used to influence intention to improve performance and facilitate behavior change and maintenance[98]. The interpretation of scores from this index can serve as a reflective guide for clinicians who are viewed as active and responsible agent in their knowledge acquisition and professional development process[99]. An important body of work has found that reflection on practice can support EBP in rehabilitation by allowing clinicians to recognize practice inconsistencies, improve deliberateness and become more open to opportunities to engage in the EBP process[100–102]. With this purpose in mind, the index could be integrated into the regulatory professional portfolio for periodic self-evaluation. Such portfolios offer the opportunity for clinicians to reflect on their experience, identify their strengths and weaknesses and make sense of how they can improve their practice[101,103]. Before implementing this index as a self-reflection tool, a feasibility and acceptability study could be conducted in partnership with regulatory bodies and clinicians to determine the perceived usefulness of the index as a self-reflection tool on professional practice.

These two broad research avenues could be expanded to develop the testing and use of the PIRE-CDMI with other rehabilitation professionals including speech-language pathologists, chiropractors, respiratory therapists, psychologists, and social workers, and with other health care professionals such as nurses and physicians. The PIRE-CDMI could also be adapted to English and French-speaking individuals outside of Canada, however a thorough cultural adaptation and reassessment is required before using the index in a measurement context that differentiates from the ones reported in this study.

Concluding statement

In conclusion, this dissertation makes an original and valuable contribution to the practice, methods, and theory of EBP measurement. The five manuscripts presented in this thesis (1) highlight shortcomings of existing measures and substantiate the need for a robust, pragmatic, and multidimensional measure of EBP; (2) identify and discuss epistemic tensions inherent in measuring EBP; (3) describe the development of the first weighted multidimensional index in rehabilitation which generates a mathematically and theoretically valid total score of a rehabilitation clinician's propensity to integrate research evidence into CDM; and (4) contribute validity evidence for the interpretation of scores from this newly developed index with OTs and PTs in Canada. There is promise for the PIRE-CDMI to be used as a self-reflection tool for clinicians to help support the uptake of research evidence into clinical decision making.

Robust EBP measures have the potential to transform clinical practice by (1) empowering clinicians in their mission to provide the best possible care and improve patient outcomes, and (2) informing researchers and decision-makers on the individual and organizational supports needed to increase the likelihood that research evidence will be integrated into CDM. In conducting the research for this dissertation, I have demonstrated methodological rigour and originality; engaged with and learned from end-users of the index; and adopted a socially conscientious approach to the development of a new measure.

References

1. World Health Organization. Rehabilitation. Published November 2021. https://www.who.int/news-room/fact-sheets/detail/rehabilitation

2. Emparanza JI, Cabello JB, Burls AJE. Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *J Eval Clin Pract*. 2015;21(6):1059-1065. doi:10.1111/jep.12460

3. Leland NE, Crum K, Phipps S, Roberts P, Gage B. Advancing the Value and Quality of Occupational Therapy in Health Service Delivery. *Am J Occup Ther*. 2015;69(1):6901090010p1-6901090010p7. doi:10.5014/ajot.2015.691001

4. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1-7. doi:10.1186/1472-6920-5-1

5. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Evidence-Based Medicine E-Book: How to Practice and Teach EBM*. 5th edition. Elsevier Health Sciences; 2018.

6. Dijkers MP, Murphy SL, Krellman J. Evidence-Based Practice for Rehabilitation Professionals: Concepts and Controversies. *Arch Phys Med Rehabil*. 2012;93(8):S164-S176. doi:10.1016/j.apmr.2011.12.014

7. CAOT, ACOTRO, ACOTUP, PAC. Joint position statement on evidence-based occupational therapy. *Can J Occup Ther*. 2009;66(5):267-269. https://www.caot.ca/document/3697/J%20-%20Joint%20Position%20Statement%20on%20Evidence%20based%20OT.pdf

8. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implement Sci*. 2014;9(1):118. doi:10.1186/s13012-014-0118-8

9. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, de Pedro-Gómez JE. Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the

literature: Evidence-based physiotherapy instruments. *J Eval Clin Pract*. 2014;20(6):767-778. doi:10.1111/jep.12180

10. Buchanan H, Siegfried N, Jelsma J. Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review: Evidence-based Practice Survey Instruments. *Occup Ther Int*. 2016;23(2):59-90. doi:10.1002/oti.1398

11. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? *J Rehabil Med*. 2012;44(2):97-98. doi:10.2340/16501977-0938

12. BISHOP PA, HERRON RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int J Exerc Sci*. 2015;8(3):297-302.

13. Jamieson S. Likert scales: How to (ab) use them? *Med Educ*. 2004;38(12):1217-1218.

14. Mayo N. *ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*. 1st edition. International Society for Quality of Life Research; 2015.

15. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79(387):516-524.

16. Al Zoubi F, Mayo N, Rochette A, Thomas A. Applying modern measurement approaches to constructs relevant to evidence-based practice among Canadian physical and occupational therapists. *Implement Sci*. 2018;13(1):152. doi:10.1186/s13012-018-0844-4

17. Thomas A, Zoubi FA, Mayo NE, et al. Individual and organizational factors associated with evidence-based practice among physical and occupational therapy recent graduates: A cross-sectional national study. *J Eval Clin Pract*. 2020;n/a(n/a). doi:https://doi.org/10.1111/jep.13518

18. Brazier J, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32). doi:10.3310/hta16320

19. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The First Stage of Developing Preference-Based Measures: Constructing a Health-State Classification Using Rasch Analysis. *Qual Life Res*. 2009;18(2):253-265.

20. American Educational Research Association., American Psychological Association., National Council on Measurement in Education., Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for Educational and Psychological Testing*.; 2014.

21. Howell J. CBC/HB for Beginners. Sawtooth Software. Published 2009. Accessed April 20, 2022. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-for-beginners

22. Bradburn N, Cartwright N, Fuller J. Chapter 5: A Theory of Measurement. In: *Measurement in Medicine : Philosophical Essays on Assessment and Evaluation*. ; 2017. Accessed June 23, 2022. https://web-p-ebscohost-com.proxy3.library.mcgill.ca/ehost/ebookviewer/ebook/bmxlYmtfXzE2MDkyNzJfX0FO0?sid=983c7550-88d1-4ee8-89c0-5320cfb680da@redis&vid=0&format=EB&lpid=lp_73&rid=0

23. Ashcroft RE. Current epistemological problems in evidence based medicine. *J Med Ethics*. 2004;30(2):131-135. doi:10.1136/jme.2003.007039

24. O'donovan B, Price M, Rust C. Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teach High Educ*. 2004;9(3):325-335.

25. Ullman JB. Structural equation modeling: Reviewing the basics and moving forward. *J Pers Assess*. 2006;87(1):35-50.

26. Chang H. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press; 2004.

27. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018;52(6):641-653. doi:https://doi.org/10.1111/medu.13574

28. Messick S. Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educ Res*. 1989;18(2):5-11. doi:10.3102/0013189X018002005

29. Messick S. Standards of validity and the validity of standards in performance asessment. *Educ Meas Issues Pract*. 1995;14(4):5-8.

30. Picciotto R. Evaluation as a social practice: Disenchantment, rationalities and ethics. *Eval Program Plann*. 2021;87:101927. doi:10.1016/j.evalprogplan.2021.101927

31. Apóstolo J, Cooke R, Bobrowicz-Campos E, et al. Predicting risk and outcomes for frail older adults: an umbrella review of frailty screening tools. *Jbi Database Syst Rev Implement Rep*. 2017;15(4):1154-1208. doi:10.11124/JBISRIR-2016-003018

32. Malik RF, Buljac-Samardžić M, Akdemir N, Hilders C, Scheele F. What do we really assess with organisational culture tools in healthcare? An interpretive systematic umbrella review of tools in healthcare. *BMJ Open Qual*. 2020;9(1):e000826. doi:10.1136/bmjoq-2019-000826

33. Belita E, Squires JE, Yost J, Ganann R, Burnett T, Dobbins M. Measures of evidence-informed decision-making competence attributes: a psychometric systematic review. *BMC Nurs*. 2020;19(1):44. doi:10.1186/s12912-020-00436-8

34. Streiner DL, Norman GR, Cairney J. *Validity*. Oxford University Press Accessed July 24, 2022. http://oxfordmedicine.com/view/10.1093/med/9780199685219.001.0001/med-9780199685219-chapter-10

35. Messick S. Validity. In: *Educational Measurement. American Council on Education Series on Higher Education.* 3rd ed. Oryx Press; 1993.

36. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ*. 2014;19(2):233-250. doi:10.1007/s10459-013-9458-4

37. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-Enhanced Simulation to Assess Health Professionals: A Systematic Review of Validity Evidence,

Research Methods, and Reporting Quality. *Acad Med*. 2013;88(6):872-883. doi:10.1097/ACM.0b013e31828ffdcf

38. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837.

39. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ*. 2017;22(4):853-867. doi:10.1007/s10459-016-9716-3

40. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. *Qual Life Res*. 2010;19(6):907-917. doi:10.1007/s11136-010-9646-0

41. Mavranezouli I, Brazier JE, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Qual Life Res*. 2011;20(3):321-333. doi:10.1007/s11136-010-9768-4

42. Wallin L, Boström AM, Gustavsson JP. Capability Beliefs Regarding Evidence-Based Practice are Associated with Application of EBP and Research Use: Validation of a New Measure. *Worldviews Evid Based Nurs*. 2012;9(3):139-148. doi:10.1111/j.1741-6787.2012.00248.x

43. Blackman IR, Giles T. Psychometric Evaluation of a Self-Report Evidence-Based Practice Tool Using Rasch Analysis. *Worldviews Evid Based Nurs*. 2015;12(5):253-264. doi:10.1111/wvn.12105

44. Spurlock D, Wonder AH. Validity and Reliability Evidence for a New Measure: The Evidence-Based Practice Knowledge Assessment in Nursing. *J Nurs Educ*. 2015;54(11):605-613. doi:10.3928/01484834-20151016-01

45. Contreras-Huerta LS, Pisauro MA, Apps MAJ. Effort shapes social cognition and behaviour: A neuro-cognitive framework. *Neurosci Biobehav Rev*. 2020;118:426-439. doi:10.1016/j.neubiorev.2020.08.003

46. Croxson PL, Walton ME, O'Reilly JX, Behrens TE, Rushworth MF. Effort-based cost–benefit valuation and the human brain. *J Neurosci*. 2009;29(14):4531-4541.

47. Kurzban R, Duckworth A, Kable JW, Myers J. An opportunity cost model of subjective effort and task performance. *Behav Brain Sci*. 2013;36(6):661-679.

48. Marquis P, Keininger D, Acquadro C, de la Loge C. Translating and evaluating questionnaires: cultural issues for international research. *Assess Qual Life Clin Trials*. Published online 2005:77-93.

49. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, USA; 2015.

50. Tanzer NK. Developing tests for use in multiple languages and cultures: A plea for simultaneous development. *Adapt Educ Psychol Tests Cross-Cult Assess*. Published online 2005:235-263.

51. Keller SD, Ware JE, Gandek B, et al. Testing the Equivalence of Translations of Widely Used Response Choice Labels: Results from the IQOLA Project. Published online 1998:12.

52. Luo N, Li M, Chevalier J, Lloyd A, Herdman M. A comparison of the scaling properties of the English, Spanish, French, and Chinese EQ-5D descriptive systems. *Qual Life Res*. 2013;22(8):2237-2243. doi:10.1007/s11136-012-0342-0

53. Poissant L, Mayo NE, Wood-Dauphinee S, Clarke AE. The development and preliminary validation of a Preference-Based Stroke Index (PBSI). *Health Qual Life Outcomes*. Published online 2003:15.

54. Kuspinar A. *Development of a Preference-Based Measure for Multiple Sclerosis: The Preference-Based Multiple Sclerosis Index (PBMSI)*. McGill University; 2014. Accessed May 3, 2019. https://onlinelibrary.wiley.com/doi/abs/10.1111/scs.12482

55. Moga AM, Twells LK, Mayo NE. The prototype of a preference-based index of weight-related quality of life: demonstrating the possibilities. *Qual Life Res*. Published online May 24, 2022. doi:10.1007/s11136-022-03156-5

56. Glasgow RE. What Does It Mean to Be Pragmatic? Pragmatic Methods, Measures, and Models to Facilitate Research Translation. *Health Educ Behav*. 2013;40(3):257-265. doi:10.1177/1090198113486805

57. D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Published online 1998:17.

58. Estabrooks CA, Thompson DS, Lovely JJE, Hofmeyer A. A guide to knowledge translation theory. *J Contin Educ Health Prof*. 2006;26(1):25-36.

59. Graham, Tetroe, Group KTR. Some theoretical underpinnings of knowledge translation. *Acad Emerg Med*. 2007;14(11):936-941.

60. Kitson A, Harvey G, McCormack B. Enabling the implementation of evidence based practice: a conceptual framework. *BMJ Qual Saf*. 1998;7(3):149-158.

61. 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(1):50.

62. Aarons GA, Hurlburt M, Horwitz SM. Advancing a Conceptual Model of Evidence-Based Practice Implementation in Public Service Sectors. *Adm Policy Ment Health Ment Health Serv Res*. 2011;38(1):4-23. doi:10.1007/s10488-010-0327-7

63. King MF, Bruner GC. Social desirability bias: A neglected aspect of validity testing. *Psychol Mark*. 2000;17(2):79-103.

64. Fisher RJ. Social desirability bias and the validity of indirect questioning. *J Consum Res*. 1993;20(2):303-315.

65. McCready T. Portfolios and the assessment of competence in nursing: A literature review. *Int J Nurs Stud*. 2007;44(1):143-151.

66. Kleiner M, Kinsella E, Miciak M, Teachman G, McCabe E, Walton D. An integrative review of the qualities of a ‘good’ physiotherapist. *Physiother Theory Pract*. Published online December 9, 2021:1-27. doi:10.1080/09593985.2021.1999354

67. Aromataris E. Twitter. Published April 20, 2022. https://twitter.com/JBIEBHC/status/1516980469107109890

68. Akram W, Hussein MSE, Ahmad S, Mamat MN, Ismail NE. Validation of the knowledge, attitude and perceived practice of asthma instrument among community pharmacists using Rasch analysis. *Saudi Pharm J*. 2015;23(5):499-503. doi:10.1016/j.jsps.2015.01.011

69. Belvedere SL, de Morton NA. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *J Clin Epidemiol*. 2010;63(12):1287-1297.

70. Conrad KJ, Smith Jr EV. International conference on objective measurement: applications of Rasch analysis in health care. *Med Care*. Published online 2004:I1-I6.

71. Harris-Haywood S, Goode T, Gao Y, et al. Psychometric Evaluation of a Cultural Competency Assessment Instrument for Health Professionals: *Med Care*. 2014;52(2):e7-e15. doi:10.1097/MLR.0b013e31824df149

72. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value Health*. 2015;18(1):25-34. doi:10.1016/j.jval.2014.10.005

73. Sawtooth. The CBC/HB System Technical Paper V5.6. Published 2009. Accessed April 26, 2022. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-technical-paper

74. Cheung KL, Wijnen BFM, Hollin IL, et al. Using Best–Worst Scaling to Investigate Preferences in Health Care. *PharmacoEconomics*. 2016;34(12):1195-1209. doi:10.1007/s40273-016-0429-5

75. Smits PA, Champagne F, Brodeur JM. A mixed method study of propensity for participatory evaluation. *Eval Program Plann*. 2011;34(3):217-227. doi:10.1016/j.evalprogplan.2011.03.002

76. Gagliardi AR, Kothari A, Graham ID. Research agenda for integrated knowledge translation (IKT) in healthcare: what we know and do not yet know. 2017;71(2):105-106. doi:10.1136/jech-2016-207743 %J Journal of Epidemiology and Community Health

77. Kothari A, Wathen CN %J JECH. Integrated knowledge translation: digging deeper, moving forward. 2017;71(6):619-623.

78. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. *Psicol Reflex E Crítica*. 2017;30(1):3. doi:10.1186/s41155-016-0057-1

79. Raykov T. Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *Br J Math Stat Psychol*. 2008;61(2):275-285.

80. Rammstedt B, Beierlein C. Can't we make it any shorter? *J Individ Differ*. Published online 2014.

81. Ziegler M, Kemper C, Kruyen P. Short Scales - Five Misunderstandings and Ways to Overcome Them. *J Individ Differ*. 2014;35:185-189. doi:10.1027/1614-0001/a000148

82. Khadjesari Z, Boufkhed S, Vitoratou S, et al. Implementation outcome instruments for use in physical healthcare settings: a systematic review. *Implement Sci*. 2020;15(1):66. doi:10.1186/s13012-020-01027-6

83. Hitch D, Nicola-Richmond K, Richards K, Stefaniak R. Student perspectives on factors that influence the implementation of evidence-based practice in occupational therapy. *JBI Evid Implement*. 2021;19(4):409-418. doi:10.1097/XEB.0000000000000285

84. McCullough D. A user’s guide to conjoint analysis. *Mark Res*. 2002;14(2).

85. Hurley AE, Scandura TA, Schriesheim CA, et al. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *J Organ Behav*. Published online 1997:667-683.

86. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2. *Med Care*. 1996;34(7):702-722.

87. Bernhardsson S, Johansson K, Nilsen P, Öberg B, Larsson MEH. Determinants of Guideline Use in Primary Care Physical Therapy: A Cross-Sectional Survey of Attitudes, Knowledge, and Behavior. *Phys Ther*. 2014;94(3):343-354. doi:10.2522/ptj.20130147

88. Upton D, Stephens D, Williams B, Scurlock-Evans L. Occupational Therapists’ Attitudes, Knowledge, and Implementation of Evidence-Based Practice: A Systematic Review of Published Research. *Br J Occup Ther*. 2014;77:24. doi:10.4276/030802214X13887685335544

89. Scurlock-Evans L, Upton P, Upton D. Evidence-Based Practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014;100(3):208-219. doi:10.1016/j.physio.2014.03.001

90. Saunders H, Gallagher-Ford L, Kvist T, Vehviläinen-Julkunen K. Practicing Healthcare Professionals’ Evidence-Based Practice Competencies: An Overview of Systematic Reviews. *Worldviews Evid Based Nurs*. 2019;16(3):176-185. doi:https://doi.org/10.1111/wvn.12363

91. Rochette A, Brousseau M, Vachon B, Engels C, Amari F, Thomas A. What occupational therapists’ say about their competencies’ enactment, maintenance and development in practice? A two-phase mixed methods study. *BMC Med Educ*. 2020;20. doi:10.1186/s12909-020-02087-4

92. Arumugam V, MacDermid JC, Walton D, Grewal R. Attitudes, knowledge and behaviors related to evidence-based practice in health professionals involved in pain management. *JBI Evid Implement*. 2018;16(2):107-118. doi:10.1097/XEB.0000000000000131

93. Wang Y, King K, Moran B, et al. Occupational and Physical Therapists' Perception of Evidence-Based Practice. *J Allied Health Wash*. 2019;48(2):119-126,126A-126C.

94. Boström AM, Sommerfeld DK, Stenhols AW, Kiessling A. Capability beliefs on, and use of evidence-based practice among four health professional and student groups in geriatric care: A cross sectional study. *PLOS ONE*. 2018;13(2):e0192017. doi:10.1371/journal.pone.0192017

95. Hasani F, MacDermid JC, Tang A, Kho M, Alghadir AH, Anwer S. Knowledge, attitude and implementation of evidence-based practice among physiotherapists working in the kingdom of Saudi arabia: a cross-sectional survey. In: *Healthcare*. Vol 8. Multidisciplinary Digital Publishing Institute; 2020:354.

96. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci*. 2012;7(1):38.

97. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Appl Psychol*. 2008;57(4):660-680.

98. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10):S70-S81.

99. Brooks JG, Brooks MG. *In Search of Understanding: The Case for Constructivist Classrooms*. Ascd; 1999.

100. Craik J, Rappolt S. Theory of Research Utilization Enhancement: A Model for Occupational Therapy. *Can J Occup Ther*. 2003;70(5):266-275. doi:10.1177/000841740307000503

101. Vachon B, Durand MJ, Leblanc J. Empowering occupational therapists to become evidence-based work rehabilitation practitioners. *Work Read Mass*. 2010;37:119-134. doi:10.3233/WOR-2010-1063

102. Hallé MC, Mylopoulos M, Rochette A, et al. Attributes of evidence-based occupational therapists in stroke rehabilitation. *Can J Occup Ther*. 2018;85(5):351-364. doi:10.1177/0008417418802600

103. Gordon JA, Campbell CM. The role of ePortfolios in supporting continuing professional development in practice. *Med Teach*. 2013;35(4):287-294.