

SYSTEMATIC REVIEW

Open Access



Instruments measuring evidence-based practice behavior, attitudes, and self-efficacy among healthcare professionals: a systematic review of measurement properties

Nils Gunnar Landsverk^{1*} , Nina Rydland Olsen² and Therese Brovold¹

Abstract

Background Evidence-based practice (EBP) is well known to most healthcare professionals. Implementing EBP in clinical practice is a complex process that can be challenging and slow. Lack of EBP knowledge, skills, attitudes, self-efficacy, and behavior can be essential barriers that should be measured using valid and reliable instruments for the population in question. Results from previous systematic reviews show that information regarding high-quality instruments that measure EBP attitudes, behavior, and self-efficacy in various healthcare disciplines need to be improved. This systematic review aimed to summarize the measurement properties of existing instruments that measure healthcare professionals' EBP attitudes, behaviors, and self-efficacy.

Methods We included studies that reported measurement properties of instruments that measure healthcare professionals' EBP attitudes, behaviors, and self-efficacy. Medline, Embase, PsycINFO, HaPI, AMED via Ovid, and Cinahl via Ebscohost were searched in October 2020. The search was updated in December 2022. The measurement properties extracted included data on the item development process, content validity, structural validity, internal consistency, reliability, and measurement error. The quality assessment, rating of measurement properties, synthesis, and modified grading of the evidence were conducted in accordance with the COSMIN methodology for systematic reviews.

Results Thirty-four instruments that measure healthcare professionals' EBP attitudes, behaviors or self-efficacy were identified. Seventeen of the 34 were validated in two or more healthcare disciplines. Nurses were most frequently represented ($n=53$). Despite the varying quality of instrument development and content validity studies, most instruments received sufficient (+) ratings on content validity, with the quality of evidence graded as "very low" in most cases. Structural validity and internal consistency were the measurement properties most often assessed, and reliability and measurement error were most rarely assessed. The quality assessment results and overall rating of these measurement properties varied, but the quality of evidence was generally graded higher for these properties than for content validity.

Conclusions Based on the summarized results, the constructs, and the population of interest, several instruments can be recommended for use in various healthcare disciplines. However, future studies should strive to use qualitative methods to further develop existing EBP instruments and involve the target population.

*Correspondence:

Nils Gunnar Landsverk
nilsgunn@oslomet.no

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Trial registration This review is registered in PROSPERO. CRD42020196009. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020196009

Keywords Evidence-based practice, Healthcare professional, Attitudes, Self-efficacy, Behavior, Development, Instrument, Validity, Reliability, Measurement error

Contributions to the literature

- Evidence-based practice (EBP) is well-known to most healthcare professionals and has become the gold standard in healthcare. To implement EBP sufficiently among healthcare personnel, we need valid and reliable instruments to measure EBP attitudes, self-efficacy, and behavior.
 - This study identified several instruments that can be recommended in different healthcare disciplines, adding knowledge that could help choose an instrument for use in future studies implementing EBP in clinical practice.
 - Our findings also underpin the importance of involving the target population and using qualitative methods when developing new EBP instruments or adapting existing EBP instruments measuring EBP attitudes, self-efficacy, and behavior.

Background

Evidence-based practice (EBP) is well known to most healthcare professionals. EBP refers to the integration of the best available research evidence with clinical expertise and patient characteristics and preferences [1]. EBP has become the gold standard in healthcare. Implementing EBP in clinical practice is associated with high-quality care, such as improved patient outcomes, reduced costs, and increased job satisfaction [2–6].

Implementing EBP in clinical practice is a complex process that is challenging and slow [3, 7]. The implementation of EBP can be hindered by barriers, including organizational, cultural, or clinician-related factors. At clinician-related level, research shows that a lack of EBP knowledge, insufficient skills, negative attitudes, low self-efficacy, and lack of EBP behaviors can be essential barriers [8, 9]. The different steps of the EBP process require that healthcare professionals understand the concepts of EBP (knowledge) and have the practical skills to do EBP activities, such as searching electronic databases or using critical appraisal tools (skills) [1, 10]. Further, the healthcare professionals' confidence in their ability to perform EBP activities (self-efficacy), and their beliefs in the positive benefits of EBP (attitudes), are known to be associated with the likelihood of EBP being successfully implemented in clinical practice (behavior) [10–12].

Strategies to improve EBP implementation should be tailored based on the healthcare professionals' perceived barriers [13–15]. However, many healthcare institutions are unaware of potential barriers that could be related to EBP knowledge, skills, attitudes, self-efficacy, and behavior among their workers [7]. These EBP constructs should be measured using valid and reliable instruments for the population in question [10]. Former systematic reviews have recommended using and further developing instruments such as the Fresno test as a measure of EBP knowledge and skills across healthcare disciplines based on existing documentation of validity and reliability on this instrument [7, 10, 16–19]. However, such clear recommendations do not exist for instruments that measure EBP attitudes, self-efficacy, and behavior.

Although several reviews have assessed instruments that measure EBP attitudes, behavior or self-efficacy [20–25], none focused on all three constructs, nor did they include studies across different healthcare disciplines. For instance, Hoegen et al. [20] included only self-efficacy instruments, and Oude Rengerink et al. [21] included only instruments measuring EBP behavior. The reviews from Belita et al. [25], Hoegen et al. [20], Leung et al. [22], Fernández-Domínguez et al. [24], and Buchanan et al. [23] included studies from one specific healthcare discipline only. A review focusing on all three constructs are needed, given the known associations between these constructs [10–12]. In addition, including studies across different healthcare disciplines could make the review more relevant for researchers targeting an interdisciplinary population.

Methodological limitations across several previous reviews may influence whether one can trust existing recommendations. Although most of the reviews evaluated the included instruments' measurement properties [20, 22–25], only Hoegen et al. [20] and Buchanan et al. [23] assessed the risk of bias in the studies included. In addition, none of the reviews rated the quality of the instruments' development processes in detail [26], and only Hoegen et al. [20] graded the quality of the total body of evidence per instrument using a modified GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach.

In short, the results from previous systematic reviews show that information regarding high-quality instruments that measure EBP attitudes, behavior, and self-efficacy

among various healthcare disciplines is still lacking. A methodologically sound review is needed to evaluate whether instruments that measure EBP attitudes, behavior, and self-efficacy can be recommended across different healthcare disciplines.

Objectives

This systematic review aimed to summarize the measurement properties of existing instruments that measure healthcare professionals' EBP attitudes, behaviors, and self-efficacy. We aimed to review the included studies' methodological quality systematically and to evaluate the instruments' development process, content validity, structural validity, internal consistency, reliability, and measurement error in accordance with the Consensus-based standards for the selection of health measurement instruments (COSMIN) methodology for systematic reviews [26–28].

Methods

This systematic review was conducted and reported following the PRISMA 2020 checklist (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [29]. The checklist is presented in Additional file 5.

Eligibility criteria

Studies were included if they met the following criteria: included healthcare professionals (e.g., nurses, physiotherapists, occupational therapists, medical doctors, psychologists, dentists, pharmacists, social workers) from primary or specialized healthcare; reported findings from the development of or the validation process of self-reported EBP instruments; described instruments measuring EBP attitudes, behavior or self-efficacy, or a combination of these EBP constructs; used a quantitative or qualitative design; and published in English or a Scandinavian language.

Studies were excluded based on the following criteria: included undergraduate students or samples from school setting; did not present any psychometric properties; focused on evidence-based diagnosis or management rather than on EBP in general; focused on the effect of implementation strategies rather than on the development or validation of an instrument; and described instruments measuring only EBP knowledge or skills.

Information sources

The following databases were included in two searches conducted in October 2020 and December 2022: MEDLINE, Embase, PsycINFO, HaPI, and AMED via Ovid, Cinahl via Ebscohost, Web of Science, and Google Scholar. In addition, we used other sources to supplement the search in the electronic databases, including

searches in the reference lists of included studies and searches for gray literature. The gray literature search included targeted website searches, advanced Google searches, gray literature databases and catalogs of gray literature, and searches for theses, dissertations, and conference proceedings. The search strategy is described in Additional file 1.

Search strategy

The search strategy was developed in consultation with and conducted by two academic librarians from OsloMet University Library. The search included terms that were related to or described the nature of the objectives and the inclusion criteria and were built around the following five elements: (1) evidence-based practice, (2) health personnel, (3) measurement and instruments, (4) psychometrics, and (5) behavior, attitude, self-efficacy.

Selection process

Titles and abstracts of studies retrieved in the search were screened independently by two review team members (NGL and TB). The studies that potentially met the inclusion criteria were identified, and the full texts of these studies were assessed for eligibility by two review members (NGL and TB). In cases of uncertainty regarding inclusion of studies, a third review member was consulted to reach a consensus (NRO). The screening and full-text assessment were conducted using Covidence systematic review software [30].

Data extraction

Data extraction was piloted on four references using a standard form completed by the first author and checked by two other review members (NRO and TB). The following data on study characteristics were extracted: author(s), publication year, title, aim, study country, study design, sample size, response rate, population/healthcare discipline description, and study setting. Data on the instruments were also extracted, including instrument name, EBP constructs measured (EBP attitudes, behaviors, and self-efficacy), theoretical framework used, EBP steps covered (ask, search, appraise, integrate, evaluate), number of items, number of subscales, scale type, instrument language, availability of questions, and translation procedure. Data on the EBP constructs measured were based on definitions from the CREATE framework (Classification Rubric for Evidence-Based Practice Assessment Tools in Education) [10]. In line with the CREATE framework, we defined the EBP constructs as follows: (1) EBP attitudes: the values ascribed to the importance and usefulness of EBP in clinical decision-making, (2) EBP self-efficacy: the judgment regarding one's ability to perform a specific EBP activity, and (3) EBP behavior: what is being

done in practice. Finally, data on the instrument's measurement properties were extracted, including data on the item development process, content validity, structural validity, internal consistency, reliability, and measurement error. Data extraction on all items was performed by the first author.

Study quality assessment

The review members (NGL, TB, and NRO) independently assessed the methodological quality of each study, using the COSMIN risk of bias checklist for systematic reviews of self-reported outcome measures [27]. Two members reviewed each study. The COSMIN checklist contains standards referring to the quality of each measurement property of interest in this review [27, 31]. The review members followed COSMIN's four-point rating system, rating the standard of each property as "very good," "adequate," "doubtful," or "inadequate" [27]. The lowest rating per measurement property was used to determine the risk of bias on that particular property, following the "worst score counts" principle [32]. After all the studies were assessed separately by the review members, a consensus on the risk of bias ratings was reached in face-to-face meetings.

Synthesis methods

The evidence synthesis process was conducted using the COSMIN methodology [26, 31]. The review members rated all the results separately, and a consensus was reached in face-to-face meetings. Instrument development and content validity studies were rated independently by the review authors according to criteria determining whether the instrument's items adequately reflected the construct to be measured [26]. These included five criteria on relevance, one criterion on comprehensiveness, and four criteria on comprehensibility [26]. The relevance, comprehensiveness, and comprehensibility per study were rated as *sufficient* (+), *insufficient* (-), *inconsistent* (+/-) or *indeterminate* (?). The reviewers also rated the instruments themselves. An overall rating was given for the relevance, comprehensibility, and comprehensiveness of each instrument, combining the results from the ratings of each study with the reviewers' ratings on the same instrument. The overall rating could not be *indeterminate* (?) because the reviewers' ratings were always available [26]. The assessment of instrument development studies included evaluating the methods used to generate items (concept elicitation) and the methods used to test the new instrument [26]. COSMIN recommends using qualitative methods, involving the target population, when developing instrument items [26].

Results for structural validity, internal consistency, reliability, and measurement error were rated independently against the COSMIN criteria for good measurement properties [28, 33, 34]. Each measurement property was rated as *sufficient* (+), *insufficient* (-) or *indeterminate* (?). To conclude each instrument, an overall rating was given for each instrument per property by jointly assessing the results from all the available studies. If the results per property per instrument were consistent, the results could be qualitatively summarized and rated overall as *sufficient* (+), *insufficient* (-), *inconsistent* (+/-) or *indeterminate* (?). More information on the COSMIN criteria for good measurement properties is provided in Additional file 2. Details on the COSMIN guideline for assessing and calculating structural validity, internal consistency, reliability, and measurement error can be found elsewhere [28, 31].

Certainty assessment

After rating the summarized results per instrument per property against the criteria for good measurement properties, we graded the quality of this evidence to indicate whether or not the overall ratings were trustworthy. The GRADE approach is used to grade the quality of evidence on four levels: *high*, *moderate*, *low*, and *very low* [35]. We used the COSMIN's modified GRADE approach, where four of the five original GRADE factors are adopted for grading the quality of evidence in systematic reviews of patient-reported outcome measures [28]. We downgraded the quality of evidence when there was concern about the results related to any of these four factors: *risk of bias*, *inconsistency*, *imprecision* or *indirectness*. Further details on the modified GRADE approach are provided in "COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs)—user manual" [28]. The quality of evidence was not graded in cases where the overall rating for a measurement property was indeterminate (?) [28]. Nor was evidence graded in cases where the overall ratings were inconsistent and impossible to summarize [31].

Results

Study selection

The search strategy identified 9405 studies. Five thousand five hundred and forty-two studies were screened for eligibility, and 156 were assessed in full text. Seventy-five studies were selected for inclusion. In addition, two studies were included via a search in gray literature. A total of 77 studies were included in the review. The PRISMA flow diagram is presented in Fig. 1.

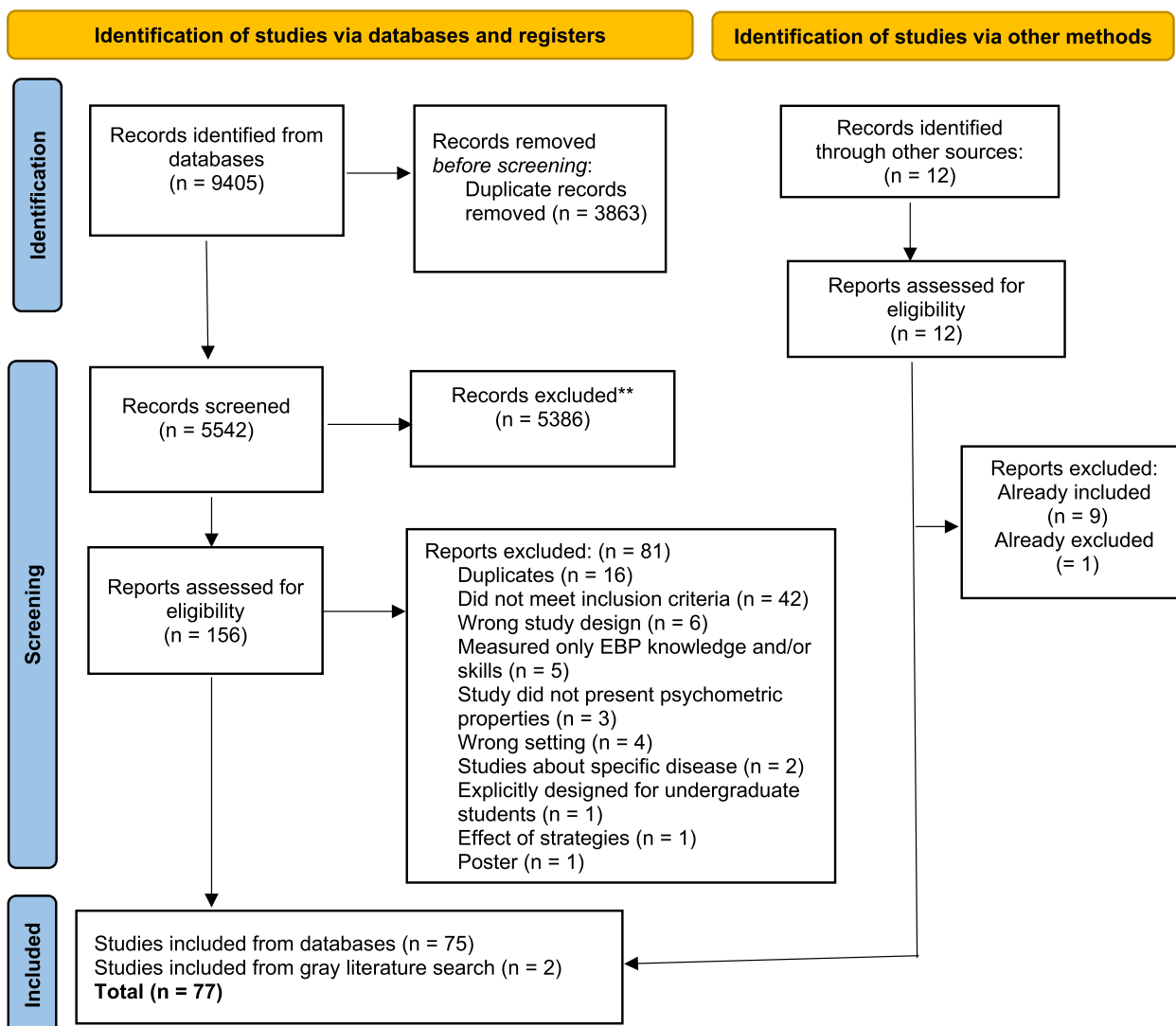


Fig. 1 PRISMA flow diagram of the selection process

Study characteristics

The 77 included studies [36–111] comprised 34 instruments measuring EBP attitudes, behavior or self-efficacy, alone or combined. Twenty-four instruments measured EBP attitudes, 21 measured behavior, and 16 measured EBP self-efficacy. Most instruments were multidimensional and included different subscales ($n = 25$). Eight instruments were unidimensional, and two had indeterminate dimensionality. Nurses were most frequently represented in the included studies ($n = 53$), followed by physiotherapists ($n = 19$), occupational therapists ($n = 10$), medical doctors ($n = 14$), mental health workers ($n = 16$), and social workers ($n = 7$). Ten of the included instruments had been validated in three or

more healthcare disciplines [36, 45, 56, 66, 68, 81, 85, 89, 111]. Seven instruments had been validated in two healthcare disciplines [47, 62, 63, 73, 75, 76, 82] and 17 had been validated in only one discipline [48, 64, 65, 71, 78–80, 87, 93, 95, 96, 102, 105, 107, 109, 110]. Details of the included studies and participants are presented in Additional file 3.

Quality assessment and results of development and content validity studies

Of the 77 studies included, 33 focused on instrument development and 18 focused on content validity on already developed instruments. Table 1 summarizes the quality assessment, rating, and quality of evidence on the development and content validity per instrument.

Table 1 Summarized results on quality assessment, rating, and quality of evidence on the development and content validity per instrument

Instrument (ref)	Development		Content validity (COSMIN box 1 and 2)					
	Concept elicitation (COSMIN box 1a)		Relevance		Comprehensiveness		Comprehensibility	
	COSMIN Quality rating	Sample involved?	Rating of results	Quality of evidence	Rating of results	Quality of evidence	Rating of results	Quality of evidence
EBPAS [36, 41, 97–99]	Doubtful	Yes	(+)	Very low*	(–)	Very Low*	(+)	Moderate (rob)
EBPAS-50 [45]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBPAS-36 [47, 100]	Doubtful	Yes	(+)	Low (rob)	(–)	Very low *	(+)	Moderate (rob)
EBPQ [48, 49, 51, 52, 54, 101]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Low (incon + rob)
EBP Beliefs [56, 58, 59, 99]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Moderate (rob)
EBP Beliefs-Short [102]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBP Implement [56, 58]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Moderate (rob)
EBP Implement-Short [102]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Very low *
Ethiopian EBP Implement [111]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Low (rob×2)
Al Zoubi Questionnaire [62]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBPP-S [63]	Inadequate	No	(±)	Very low *	(–)	Very low *	--	--
Jette [64, 103]	Inadequate	No	(+)	Very low *	(+)	Moderate (rob)	(+)	Moderate (rob)
Bernhardsson [65]	Inadequate	No	(+)	Very low *	(+)	Low (rob×2)	(+)	Low (rob×2)
EBP inventory [66, 67]	Doubtful	Yes	(+)	Low (rob×2)	(+)	Very low *	(+)	Moderate (rob)
EPIC [68]	Doubtful	No	(+)	Low (rob×2)	(+)	Low (rob×2)	(+)	Low (rob×2)
MPAS [71]	Inadequate	No	(+)	Very low *	(–)	Very low *	(+)	Very low *
EBPPAS [73]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Very low *
SE-EBP [76]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBPSE [78]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBP Capability Beliefs [79]	Inadequate	No	(+)	Very low *	(+)	Very low *	(+)	Very low *
HEAT [80]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBP-KABQ [81]	Doubtful	No	(+)	Very low *	(+)	Very low *	(+)	Very low *
Quick EBP-VIK [82, 84]	Doubtful	No	(+)	Very low *	(+)	Very low *	(+)	Moderate (rob)
HS-EBP [85]	Adequate	Yes	(+)	Very low *	(+)	Very low *	(+)	Very low *
EBPRS [87, 88]	Inadequate	No	(±)	Very low *	(–)	Very low *	(+)	Very low *
EBP2 [89, 90, 92]	Doubtful	No	(+)	Very low *	(+)	Very low *	(+)	Moderate (rob)
ISP-D [93]	Adequate	Yes	(+)	Low (rob×2)	(+)	Very low *	(+)	Low (rob×2)
EBNAQ [95]	Doubtful	Yes	(+)	Low (rob×2)	(+)	Very low *	(+)	Low (rob×2)
Diermayr [96]	Doubtful	Yes	(+)	Very low *	(–)	Very low *	(+)	Very low *
EBP-COQ Prof [105, 106]	Doubtful	Yes	(+)	Low (rob×2)	(+)	Very low *	(+)	Moderate (rob)
EIDM competence measure [107]	Adequate	Yes	(+)	Moderate (rob)	(+)	Very low *	(+)	Moderate (rob)
I-SABE [108]	Doubtful	Yes	(+)	Low (rob×2)	(+)	Very low *	(+)	Low (rob×2)
Noor EBM [109]	Doubtful	Yes	(+)	Very low *	(+)	Very low *	(+)	Low (rob×2)
EBP-CBFRI [110]	Doubtful	No	(+)	Very low *	(+)	Very low *	(+)	Very low *

Overall rating of results: (+) = sufficient; (–) = insufficient; (?) = indeterminate; (±) = inconsistent

Quality assessment: VG = very good; A = adequate; D = doubtful; I = inadequate

Quality of evidence: Modified GRADE approach [28, 31]. Quality levels: high, moderate, low, and very low

Reasons for downgrade: risk of bias = “RoB”, Inconsistency = “Incon”, Imprecision = “Impre”, Indirectness = “Indir”

“--”: No grade due to lack of questionnaire access

* When based only on reviewer’s rating, i.e., not enough evidence from **or** inadequate quality of development study and not enough evidence from **or** inadequate quality of content validity study

The quality of concept elicitation (development of items) was rated as “adequate” in three studies [85, 93, 107], where a clearly reported and appropriate method

was used and a sample representing the target population was involved. A further 19 studies received a “doubtful” quality rating [36, 45, 47, 48, 62, 66, 68, 76, 78, 80–82, 89,

95, 96, 105, 108–110]. Some of these studies used qualitative methods to generate items, but the method, or parts of it, was not clearly described. In other studies, it was doubtful whether the included sample was representative of the target population, and some used quantitative methods. Some studies were rated as “doubtful” if it was stated that authors of these studies had talked or discussed the items with relevant healthcare professionals as a part of concept elicitation, but it was doubtful whether this method was suitable. Finally, 12 studies received an “inadequate” quality rating for concept elicitation [56, 63–65, 71, 73, 79, 87, 102, 111]. In these cases, it was clear that no qualitative methods that involved members of the target population were used when generating items. The item generation was usually based on theory, research, or existing instruments.

Content validity was also assessed as part of the development studies with cognitive interviews or pilot tests or in separate content validity studies performed after the instrument was developed, primarily studies translating an instrument. Some development studies assessed comprehensibility [47, 56, 65, 68, 73, 76, 78, 82, 87, 89, 93, 95, 105, 107–111] or comprehensiveness [65, 68] with interviews or pilot tests on samples representing the target population. These were rated as either “adequate” [93, 107] or “doubtful” quality [47, 56, 65, 68, 73, 76, 78, 82, 87, 89, 95, 105, 108–111]. The rest of the development studies could not be rated, either because it was unclear whether a pilot test or interview was performed, or which aspect of content validity was assessed. Most of the content validity studies assessed comprehensibility [41, 49, 51, 52, 54, 58, 59, 84, 88, 90, 92, 97–101, 103, 106] and only a few assessed relevance or comprehensiveness [59, 84, 88, 99, 103]. All content validity studies were rated as doubtful quality [41, 49, 51, 52, 54, 58, 59, 84, 88, 90, 92, 97–101, 103, 106].

Results of synthesis and certainty of evidence on content validity

With the combined results from each study’s ratings of relevance, comprehensiveness, and comprehensibility and the reviewers’ ratings, each instrument was given an overall rating (Table 1). Most instruments were rated as sufficient (+) on relevance and comprehensibility, and only 6 out of 34 instruments were rated as insufficient (–) on comprehensiveness. The quality of evidence was graded as “very low” in most cases, primarily due to no content validity studies (or inadequate quality) and not enough evidence from (or inadequate quality of) the development studies. The overall grade was, in these cases, based solely on the reviewers’ ratings and was therefore downgraded to “very low” [26].

Seven instruments (EBPAS-36, EBP Inventory, EPIC, ISP-D, EBNAQ, EBP-COQ Prof, and I-SABE) had “low”

quality evidence of sufficient “relevance” from concept elicitation studies of doubtful quality [26]. One instrument (EIDM competence measure) had “moderate” quality evidence of sufficient “relevance” from a development study of adequate quality. Two instruments (EPIC and Bernhardsson) had “low”, and another (Jette) had “moderate” quality evidence of sufficient “comprehensiveness” from a development study of doubtful quality and a content validity study of doubtful quality [26].

Ten instruments (EBPAS, EBPAS 36, EBP inventory, EBP Beliefs, EBP Implement, Jette, Quick EBP VIK, EBP2, EBP-COQ Prof, and EIDM competence measure) had “moderate” quality evidence of sufficient “comprehensibility” from content validity studies of doubtful quality or development studies of adequate quality [26]. In addition, eight instruments (EBPQ, EPIC, Bernhardsson, ISP-D, EBNAQ, I-SABE, Noor EBM, and Ethiopian EBP Implement) had “low” quality evidence of sufficient “comprehensibility” from development studies of doubtful quality or content validity studies of doubtful quality but with inconsistent results [26].

Quality assessment and results of structural validity and internal consistency studies

Structural validity was assessed in 63 studies and internal consistency in 69 studies. The quality assessment and results of rating of structural validity and internal consistency per study are presented in detail in Additional file 4.

To test structural validity, most studies used exploratory factor analyses (EFA) ($n=26$) or confirmatory factor analyses (CFA) ($n=34$), and two studies used IRT/Rasch analyses. Since CFA is preferred over EFA in the COSMIN methodology [31], only the results of CFA were rated in studies where both EFA and CFA were conducted. The quality of structural validity testing was rated as “very good” in 33 studies [36–38, 40, 42–44, 47, 49, 50, 53, 55, 72, 74, 75, 77, 79–81, 84, 86, 88, 90, 92, 94, 97–100, 105, 106, 110], “adequate” in 19 studies [39, 45, 48, 51, 52, 57, 58, 60, 62, 69, 76, 89, 91, 95, 108, 109, 111], “doubtful” in 9 studies [46, 56, 59, 61, 63, 83, 102], and as “inadequate” in two studies [66, 73]. In both cases inadequate ratings were given due to low sample sizes [31].

To test internal consistency of the items, most studies calculated and reported a Cronbach’s alpha ($n=67$), and two studies calculated and reported a person separation index. The quality of internal consistency calculations was rated as “very good” in 64 studies [36–39, 41–45, 47–63, 66, 67, 69, 71–81, 83, 84, 86, 88–92, 94, 95, 97, 99–102, 104–106, 108, 110] and as “inadequate” in five studies [46, 60, 98, 109, 111]. Inadequate ratings were given when a Cronbach’s alpha was not reported for each unidimensional subscale in a multidimensional instrument [31].

Results of synthesis and certainty of evidence of structural validity and internal consistency

Qualitatively summarized results, overall rating, and quality of evidence (COSMIN GRADE) on structural validity and internal consistency per instrument are presented in detail in Tables 2 and 3.

Eighteen instruments were rated overall as sufficient (+) structural validity (EBPAS, EBPAS-50, EBPQ, EBP Belief-single factor, EBP Implement-single factor, EBPP-S, EPIC, MPAS, HEAT, Quick EBP-VIK, HS-EBP, EBPRS, ISP-D, EBNAQ, EBP Beliefs short, EBP Implement Short, EBP-CBFRI, and Ethiopian EBP Implement), with the quality of evidence ranging from “high” to “low.” Reasons for downgrading the quality of evidence were either “risk of bias” or “inconsistency.” Six instruments were rated overall as insufficient (–) structural validity (EBP belief-multifactorial, EBP implement-multifactorial, EBPPAS-s, EBP-KABQ, EBP-COQ Prof, and I-SABE), with the quality of evidence ranging from “high” to “moderate.” The reasons for downgrading were “inconsistency” and “risk of bias.” Four instruments were rated overall as inconsistent (+/–) structural validity (EBPPAS, SE-EBP, EBP2, and EBPAS-36). In these three cases, results were inconsistent and it was not possible to give an overall rating as sufficient or insufficient (e.g., an overall rating based on the majority of studies) [31]. Finally, four instruments were rated overall as indeterminate (?) structural validity (Al Zoubi Q, EBP Inventory, EBP capability beliefs, and Noor EBM) because not all the information needed for a sufficient rating was reported [31].

Regarding internal consistency, 16 instruments were rated overall as indeterminate (?) (EBP belief-multifactorial, EBP implement-multifactorial, Al Zoubi Q, EBP Inventory, EBPPAS, EBPPAS-s, SE-EBP, EBPSE, EBP capability beliefs, EBP-KABQ, EBP2, EBP-COQ Prof, I-SABE, Noor EBM, Ethiopian EBP Implement, and EBPAS-36). Most of these instruments had Cronbach's alpha values that met the criteria for sufficient internal consistency ($\alpha > 0.70$). However, since evidence of structural validity is a prerequisite of internal consistency, they were rated as indeterminate (?) according to the COSMIN methodology [28]. Furthermore, the summarized result of internal consistency was rated and graded per subscale in cases of multifactorial instruments. This led to several instruments receiving different ratings on different subscales, such as sufficient (+), insufficient (–) or inconsistent (+/–) (EBPAS, MPAS, Quick EBP VIK, ISP-D, and EBNAQ). Seven multifactorial and five unidimensional instruments were rated as sufficient (+) on all subscales or full scales (EBPAS-50, EBPQ, EBP Beliefs-single factor, EBP Implement-single factor, EBPP-S, EPIC, HEAT, HS-EBP, EBPRS, EBP

Beliefs-Short, EBP Implement-Short, and EBP-CBFRI). The quality of evidence ranged from “high” to “low,” and the most common reason for downgrading was that the quality of evidence of structural validity on the same instrument set the starting point for the grading of internal consistency [31].

Quality assessment and results of reliability and measurement error studies

Reliability was assessed in 22 studies, and measurement error in five studies. The quality assessment and results of the rating of reliability and measurement error per study are presented in detail in Additional file 4.

To test reliability, 18 studies calculated and reported an intraclass correlation coefficient (ICC), two used Pearson's correlation, and two used the percentage of agreement. The quality of reliability testing was rated as “very good” in two studies [41, 67], “adequate” in 12 studies [39, 64, 66, 69, 83, 84, 89–92, 105, 106], “doubtful” in six studies [46, 50, 52, 54, 70, 96], and as “inadequate” in two studies [65, 103]. Reasons for a “doubtful” rating were that time intervals between measurements were longer than recommended or it was unclear whether respondents were stable between measurements or whether only Pearson's or Spearman's correlation coefficients were calculated [31]. The reason for the “inadequate” rating was that no ICC, Pearson's or Spearman's correlation coefficients were calculated [31].

To test measurement error, all studies calculated standard error of measurement (SEM), smallest (minimal) detectable change (SDC) or limits of agreement (LoA). Only one study reported information on minimal important change (MIC). The quality of measurement error testing was rated as “very good” in two studies [41, 67], “adequate” in two studies [69, 92], and as “doubtful” in one study [70]. The reason for the “doubtful” rating was that a time interval between measurements was longer than recommended.

Results of synthesis and certainty of evidence of reliability and measurement error

Qualitatively summarized results, overall rating, and quality of evidence (COSMIN GRADE) on reliability and measurement error are presented in detail in Tables 4 and 5.

The summarized result of reliability was rated and graded per subscale in cases of multifactorial instruments. This led to four instruments receiving different overall ratings on different subscales, such as sufficient (+), insufficient (–) or inconsistent (+/–) reliability (EBPAS, EBPQ, Quick EBP-VIK, and EBP2). Three instruments were rated overall as sufficient (+)

Table 2 Qualitatively summarized results, overall rating, and quality of evidence (GRADE) on *structural validity* per instrument

Instrument (ref)	Summarized result on structural validity	Rating	Quality of evidence	
			Quality level	Reason
EBPAS [36–40, 42–44, 97–99]	EFA and CFA. Four factors/subscales. Inconsistent results. Results for seven studies met the criteria (SRMR < 0.08), three studies did not (SRMR > 0.08/RMSEA > 0.06)	(+)	Moderate	– 1 incon
EBPAS-50 [45, 46]	EFA and CFA. Eight new factors. Inconsistent results, different number of factors in two studies. EFA criteria were met on eight-factor structure, no fit indices reported on CFA	(+)	Low	– 1 incon, – 1 RoB
EBPAS-36 [47, 100]	CFA showed a 36-item 12-factor scale. Inconsistent results. One study met the criteria (RMSEA = 0.052) and one study did not (RMSEA = 0.64)	(±)	No grade	
EBPQ [48–53, 55]	EFA and CFA. Three factors/subscales. Inconsistent results. Results for five studies met the criteria for CFA (SRMR < 0.08) or criteria for EFA, two studies did not	(+)	Moderate	– 1 incon
EBP Beliefs [56–60]	EFA. Disagreement about dimensionality between five studies. Results summarized in subgroups (unidimensional/multifactorial) 1. Unidimensional (single factor): criteria for EFA were met in both (two) 2. Multifactorial (four factors): inconsistent results. Criteria not met in two studies due to cross-loading, and the third for not reporting cross-loading	1: (+) 2: (–)	Moderate Moderate	– 1 RoB – 1 incon
EBP Beliefs- Short [102]	EFA. One factor. Factor loading > 0.70, Eigenvalue = 2.25	(+)	Low	– 2 RoB
EBP Implement [56, 58, 60, 61]	EFA. Disagreement about dimensionality between four studies. Results are summarized in subgroups (unidimensional/multifactorial) 1. Unidimensional (single factor): criteria for EFA were met in one study 2. Multifactorial (four/five/two factors): inconsistent results. Criteria not met in two studies due to cross-loading, and one study rated as indeterminate due to not reporting eigenvalue, total variance explained, or cross-loading	1: (+) 2: (–)	Low Moderate	– 2 RoB – 1 incon
EBP Implement-Short [102]	EFA. One factor. Factor loading > 0.85, Eigenvalue = 2.46	(+)	Low	– 2 RoB
Ethiopian EBP Implement [111]	EFA (two factors): Factor loadings > 0.40, Eigenvalues > 1, cross-loading, no cross loadings	(+)	Moderate	– 1 Rob
AI Zoubi Questionnaire [62]	IRT/Rasch: Unidimensionality: CFI, TLI, RMSEA or SRMR not reported. Local independence: items correlating > 0.3 led to removal. Monotonicity: not reported. Model fit (× 2 test): × 2 > 0.01 on three out of four subscales	(?)	No grade	
EBPP-S [63]	CFA confirmed a three-factor scale. Model fit: CFI = .96, and RMSEA = .06	(+)	Low	– 2 RoB
EBP inventory [66]	EFA. Four factors/subscales. Factor loadings > 0.30, cross-loading, tot variance explained, and eigenvalue not reported	(?)	No grade	
EPIC [69]	Unidimensional scale. EFA: all items loaded into one single factor > 0.4, tot explained variance = 71%, < 10% cross-loading	(+)	Moderate	– 1 RoB
MPAS [72]	CFA. Unidimensional model, and a modified five-item model had the best fit: US: CFA: RMSEA = 0.030, CFI 0.998 Korea: CFA: RMSEA = < 0.05, CFI > 0.95	(+) (+)	High High	
EBPPAS [73, 74]	EFA and CFA. Inconsistent result. One study met the criteria for EFA (factor loading > 0.3, eigenvalues > 1, and < 10% cross-loading). The other study did not meet the criteria for CFA (CFI = 0.90)	(±)	No grade	
EBPPAS-s [75]	CFA. Revised four-factor model, 37 items. Model fit: CFI = 0.90, RMSEA = 0.06	(–)	High	
SE-EBP [76, 77]	EFA and CFA. Three factors/subscales. Inconsistent results. One study met the criteria for EFA (factor loading > 0.3, tot explained variance = 73.01%, and < 10% cross-loading). One study did not meet the criteria for CFA (CFI = 0.91)	(±)	No grade	
EBP Capability Beliefs [79]	IRT/Rasch: Unidimensionality: CFI, TLI, RMSEA or SRMR not reported. Local independence: two small cross-loadings on the first factor; 0.23 and 0.30. Monotonicity: not reported. Model fit (× 2 test): × 2 = 42, 71	(?)	No grade	
HEAT [80]	EFA and CFA. Four-factor model. CFA model fit: SRMR = 0.063	(+)	High	
EBP-KABQ [81]	CFA: Four-factor modified model. Model fit: CFI = 0.89	(–)	High	
Quick EBP-VIK [83, 84]	EFA and CFA. Three-factor model, inconsistent results. Model fit: CFI = 0.957	(+)	Moderate	– 1 Incon
HS-EBP [86]	EFA and CFA. Five factors. Model fit: CFI = 0.99, RMSEA = 0.047, SRMR = 0.067	(+)	High	

Table 2 (continued)

Instrument (ref)	Summarized result on structural validity	Rating	Quality of evidence	
			Quality level	Reason
EBPRS [88]	EFA and CFA. Four factors. Model fit: CFI=0.96, TLI=0.96, SRMR=0.058	(+)	High	
EBP2 [89–92]	EFA and CFA. Inconsistent results. Five factors in three out of four studies. One study met the criteria for EFA (factor loading > 0.3, tot explained variance = 63%, and < 10% cross-loading), one did not report enough information to be rated, the third did not meet criteria for CFA (CFI=0.69). Overall rating is inconsistent	(±)	No grade	
ISP-D [94]	CFA. Four-factor model. Model fit: SRMR=0.075	(+)	High	
EBNAQ [95]	PCA (EFA) with Promax rotation showed three factors explaining a total variance of 54.7%, factor loadings > 0.30 and cross-loading < 10%	(+)	Moderate	– 1 RoB
EBP-COQ Prof [105, 106]	CFA. Four-factor model. Model fit: CFI=0.93 and 0.82 in two studies	(–)	High	
I-SABE [108]	EFA. Four factors. Tot variance explained = 52.6%, factor loading > 0.3, cross-loading > 10%	(–)	Moderate	– 1 RoB
Noor EBM [109]	EFA. 1. Attitude scale (five factors): factor loading = > 0.4, tot variance = 66.3%, cross-loading not reported. 2. Practice scale (two factors): factor loading = > 0.4, tot variance = 55.4%, cross-loading not reported	1: (?) 2: (?)	1,2: no grade	
EBP-CBFRI [110]	CFA. Five-factor model. Model fit: RMSEA = 0.05	(+)	High	

Overall rating of results: (+) = sufficient result; (–) = insufficient result; (?) = indeterminate result; (±) = inconsistent results

Quality of evidence: Modified GRADE-approach [28, 31]. Quality levels: high, moderate, low, and very low

Reasons for downgrade: Risk of bias = “RoB”, Inconsistency = “Incon”, Imprecision = “Impre”, Indirectness = “Indir”

CFI Comparative fit index, RMSEA Root mean square error of approximation, SRMR Standardized square residual, EFA Exploratory factor analysis, CFA Confirmatory factor analysis

reliability (EBP inventory, EPIC, and EBP-COQ Prof). The quality of evidence ranged from “high” to “low.” Reasons for downgrading the quality of evidence were either “inconsistency,” “risk of bias” or “imprecision.” Four instruments were rated overall as indeterminate (?) reliability (EBPAS-50, EBP (Jette), EBP (Bernhards-son), and EBP (Diermayr)). The reasons for indeterminate ratings were that ICC was not calculated, not reported or not reported in sufficient detail to allow rating and grading [31].

Regarding measurement error, one instrument was rated overall as sufficient (+), with the quality of evidence graded as “moderate.” It was downgraded for imprecision due to the small sample size. Since MIC was not defined, three other instruments were rated overall as indeterminate (?) measurement error [31].

Discussion

This review sought to summarize measurement properties of existing instruments that measure healthcare professionals’ EBP attitudes, behaviors, and self-efficacy. We evaluated the instruments’ development process, content validity, structural validity, internal consistency, reliability, and measurement error. Thirty-four instruments measuring EBP attitudes, behavior or self-efficacy, alone or combined, were identified.

The assessment of instrument development studies revealed that only three instruments received an “adequate” quality rating on concept elicitation (HS-EBP, ISP-D, and EIDM competence measure) [85, 93, 107]. The rest were rated “doubtful” or “inadequate.” Reasons for “doubtful” ratings were mainly related to the quality of the qualitative methods used to generate items and “inadequate” ratings were given when no qualitative methods seemed to have been used. The use of well-designed qualitative methods when constructing the items is emphasized in the updated COSMIN methodology (2018) that was used in this review [26]. However, over two-thirds of the development studies included in this review were published before the updated COSMIN methodology was published in 2018 [26]. Thus, assessing instrument development studies based on a detailed and standardized methodology to which the developers did not have access when developing instruments can be somewhat strict. At the same time, the quality of the development process (concept elicitation) has not, to our knowledge, been rated in detail in previous reviews of EBP instruments [20–25]. Thus, our findings underpin the importance that future instrument development studies should involve the target population using qualitative methods to generate items for an EBP instrument.

Table 3 Qualitatively summarized results, overall rating, and quality of evidence (GRADE) on *internal consistency* per instrument

Instrument (reference)	Summarized result on internal consistency (overall rating)	Rating	Quality of evidence	
			Quality level	Reason
EBPAS [36–39, 41–44, 97, 99]	Cronbach's alpha range per subscale: 1. Requirements 0.88–0.94, 2. Appeal 0.72–0.80, 3. Openness 0.73–0.84, 4. Divergence 0.51–0.85	1–3: (+) 4: (–)	Moderate Moderate	– 1 struc val – 1 struc val
EBPAS-50 [45, 46]	Cronbach's alpha ranged from 0.77 to 0.92 on the eight new subscales	(+)	Low	– 2 struc val
EBPAS-36 [47, 100]	Cronbach's alphas range per subscale: 1. Appeal, Fit, Balance and Divergence = 0.61–0.69, 0.62–0.68, 0.64–0.65, and 0.66–0.68. 2. All other subscales > 0.70. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
EBPQ [48–55, 101]	Cronbach's alpha range per subscale: 1. Attitudes: 0.68–0.83, 2. Practice: 0.74–0.93, 3. Knowledge/skills: 0.90–0.96	1: (+) 2,3: (+)	Low Moderate	– 1 struc val, – 1 incon – 1 struc val
EBP Beliefs [56–60]	Disagreement about dimensionality between give studies. Results are summarized in subgroups (unidimensional/multifactorial) 1. Unidimensional (single factor): Cronbach's alpha range = 0.86–0.90 2. Multifactorial (four factors): Indeterminate rating (?) since criteria for sufficient structural validity not met	1: (+) 2: (?)	Moderate No grade	– 1 struc val
EBP Beliefs- Short [102]	Cronbach's alpha on scale = 0.89	(+)	Low	– 2 Struc val
EBP Implement [56, 58, 60, 61]	Disagreement about dimensionality between four studies. Results are summarized in subgroups (unidimensional/multifactorial) 1. Unidimensional (single factor): Cronbach's alpha = 0.96 2. Multifactorial (four/five/two factors): Indeterminate rating (?) since criteria for sufficient structural validity not met	1: (+) 2: (?)	Low No grade	– 2 struc val
EBP Implement- Short [102]	Cronbach's alpha on scale = 0.81	(+)	Low	– 2 Struc val
Ethiopian EBP Implement [111]	Cronbach's alpha not reported for each subscale. total scale = 0.83	(?)	No grade	
AI Zoubi Questionnaire [62]	Person separation index per subscale: Attitudes = 0.63, Self-efficacy = 0.80, Knowledge = 0.81, Resources = 0.86. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
EBPP-S [63]	Cronbach's alpha per subscale: 1. Attitudes = 0.89, 2. Knowledge = 0.90, 3. Behavior = 0.83	1–3 = (+)	Low	– 2 struc val
EBP inventory [66, 67]	Cronbach's alpha range per subscale: 1. Decision making = 0.60–0.64, 2. All other subscales > 0.70. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
EPIC [69]	Cronbach's alpha on scale: 0.89	(+)	Low	– 1 Struc val, – 1 RoB
MPAS [71, 72]	Cronbach's alpha range per sample: 1. USA = 0.78–0.80, 2. Korea: 0.65	1: (+) 2: (–)	High Low	– 1 RoB
EBPPAS [73, 74]	Cronbach's alpha range per subscale: Familiarity = 0.91–0.92, Attitudes = 0.83–0.90, Feasibility = 0.57–0.63, Intentions = 0.80–0.86, Currently engaged = 0.86–0.87. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
EBPPAS-s [75]	Cronbach's alpha per subscale: Familiarity = 0.93, Attitudes = 0.92, Feasibility = 0.74, Currently engaged = 0.91. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
SE-EBP [76, 77]	Cronbach's alpha range per subscale: identifying the clinical problem = 0.89–0.91, searching for evidence = 0.90–0.96, implementing evidence into practice = 0.93–0.96. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
EBPSE [78]	Cronbach's alpha = 0.95. Indeterminate rating (?) since no evidence of unidimensionality	(?)	No grade	

Table 3 (continued)

Instrument (reference)	Summarized result on internal consistency (overall rating)	Rating	Quality of evidence	
			Quality level	Reason
EBP Capability Beliefs [79]	Person separation index = 0.92. Indeterminate rating (?) since no evidence of unidimensionality	(?)	No grade	
HEAT [80]	Cronbach's alpha per subscale: EBP frequency = 0.82, EBP ability = 0.89, EBP desire = 0.92, EBP barriers = 0.80	(+)	High	
EBP-KABQ [81]	Cronbach's alpha per subscale: Knowledge = 0.66, Attitude = 0.75, Behavior = 0.77, Outcome = 0.83. Indeterminate rating (?) since criteria for sufficient structural validity not met	(?)	No grade	
Quick EBP-VIK [83, 84]	Cronbach's alpha range per subscale: 1. Value = 0.78–0.89, 2. Knowledge = 0.92–0.93, 3. Implement = 0.66–0.91	1–2: (+) 3: (±)	Moderate No grade	–2: Struc val
HS-EBP [86]	Cronbach's alpha per subscale: Beliefs and attitudes = 0.93, Results from scientific research = 0.96, Development of professional practice = 0.84, Assessment of results = 0.94, Barriers or Facilitators = 0.91	(+)	High	
EBPRS [88]	Cronbach's alpha per subscale: Informational needs: 0.85, EBP knowledge: 0.78, EBP attitude: 0.87, Workplace culture: 0.84	(+)	High	
EBP2 [89–92, 104]	Cronbach's alpha range per subscale (five-factor model): Relevance: 0.91–0.94, Sympathy: 0.66–0.80, Terminology: 0.94–0.98, Practice: 0.85–0.923, Confidence: 0.93–0.94. Indeterminate (?) rating since criteria for sufficient structural validity not met	(?)	No grade	
ISP-D [94]	Cronbach's alpha per subscale (four-factor model): 1. Attitudes = 0.75, 2. Subjective norms = 0.72, 3. Perceived behavioral control = 0.63, 4. Behavioral intention = 0.84	1, 2, 4: (+) 3: (–)	1, 2, 4: High 3: High	
EBNAQ [95]	Cronbach's alpha per subscale (three-factor model): 1. Beliefs and expectations = 0.86, 2. Intention of conduct = 0.63, 3. Feelings toward the EBN = 0.70	1, 3: (+) 2: (–)	Moderate Moderate	–1 struc val –1 struc val
EBP-COQ Prof [105, 106]	Cronbach's alpha per subscale: 1. Attitude toward EBP = 0.888, 2. EBP knowledge = 0.948, 3. EBP skills = 0.817, 4. EBP utilization = 0.840	(?)	No grade	
I-SABE [108]	Cronbach's alpha per subscale: 1. EBP self-efficacy = 0.76, 2. Behavior = 0.30, 3. Attitudes = 0.64, 4. Results = 0.84. Indeterminate (?) rating since criteria for sufficient structural validity not met	(?)	No grade	
Noor EBM [109]	Cronbach's alpha on total scales (not per subscale): 1. Attitude scale = 0.81, 2. Practice scale = 0.84. Indeterminate (?) rating since criteria for sufficient structural validity not met	(?)	No grade	
EBP-CBFRI [110]	Cronbach's alpha range on subscales = 0.92–0.95	(+)	High	

Overall rating of results: (+) = sufficient result; (–) = insufficient result; (?) = indeterminate result; (±) = inconsistent results

Quality of evidence: modified GRADE approach [28, 31]. Quality levels: high, moderate, low, and very low

Reasons for downgrade: risk of bias = "RoB", Inconsistency = "Incon", imprecision = "Impre", indirectness = "Indir"

α = Cronbach's alpha

The summarized results on internal consistency showed that several instruments were rated overall as indeterminate (?) despite meeting the criteria for a sufficient (+) rating (Cronbach's alpha > 0.70). Although measuring "how well items correlate," Cronbach's alpha is often misinterpreted as a measure of the dimensionality of a scale. Whether the scores on a scale reflect the dimensionality of the construct measured is defined as structural validity and is most often assessed by factor

analysis ([112], p. 169–170, [113]). Evidence of unidimensionality of a scale or subscale is an assumption that needs to be verified before calculating Cronbach's alpha to assess the interrelatedness of the items [113]. Though internal consistency helps assess whether items on a scale or subscale are related, evidence of structural validity must come first to ensure that the interrelated items are on a scale or subscale that also reflects the construct's dimensionality. The rating of internal

Table 4 Qualitatively summarized results, overall rating, and quality of evidence (GRADE) on *reliability* per instrument

Instrument (reference(s))	Summarized result on reliability	Overall rating	Quality of evidence	
			Quality level	Reason
EBPAS [39, 41]	ICC range per subscale: 1. Requirements 0.55–0.80, 2. Appeal 0.40–0.56, 3. Openness 0.64–0.71, 4. Divergence 0.48–0.74	1, 3, 4: (±) 2: (–)	No grade High	
EBPAS-50 [46]	Pearson’s correlation total scale = 0.344. Indeterminate rating (?) since ICC not calculated	(?)	No grade	
EBPQ [50, 52, 54]	ICC range per subscale: 1. Attitudes: 0.44–0.86, 2. Practice: 0.78–0.84, 3. Knowledge/skills: 0.70–0.86	1: (±) 2, 3: (+)	No grade Moderate	– 1 RoB
Jette [64, 103]	ICC ranged from 0.37 to 0.90 (50% > 0.70). Indeterminate rating (?) since ICC per subscale not reported	(?)	No grade	
Bernhardsson [65]	Indeterminate rating (?) since no ICC or Pearson’s or Spearman’s correlations reported	(?)	No grade	
EBP inventory [66, 67]	ICC range per subscale: 1. Decision making = 0.71–0.78, 2. Subjective norm = 0.63–0.86, 3. Attitude = 0.52–0.82, 4. Perceived behavioral control = 0.80–0.83, 5. Intention and behavior = 0.76–0.86	1, 4, 5: (+) 2, 3: (+)	High Moderate	– 1 Incon
EPIC [69, 70]	ICC range = 0.89–0.92	(+)	Moderate	– 1 RoB
Quick EBP-VIK [83, 84]	ICC range per subscale: 1. Value = 0.51–0.57, 2. Knowledge = 0.70–0.88, 3. Implement = 0.63–0.84	1: (–) 2: (+) 3: (±)	High High No grade	
EBP2 [89–92]	ICC range per subscale (five-factor model). 1. Relevance = 0.69–0.94, 2. Terminology = 0.79–0.94, 3. Confidence = 0.76–0.95, 4. Practice = 0.45–0.92, 5. Sympathy = 0.47–0.77	1, 4: (+) 2, 3: (+) 5: (±)	Moderate High No grade	Incon
EBP Diermayr [96]	Mean ICC 0.67 (range 0.40–0.89). Indeterminate rating (?) since ICC per subscale not reported	(?)	No grade	
EBP-COQ Prof [105, 106]	ICC per subscale (four-factor model): 1. Attitude toward EBP = 0.840, 2. EBP knowledge = 0.966, 3. EBP skills = 0.815, 4. EBP utilization = 0.876	(+)	Moderate	– 1 impre

ICC = Intraclass correlation coefficient

Rating of results: (+) = sufficient result; (–) = insufficient result; (?) = indeterminate result; (±) = inconsistent results

Quality of evidence: modified GRADE approach [28, 31]. Quality levels: high, moderate, low, and very low

Reasons for downgrade: risk of bias = “RoB”, inconsistency = “Incon”, imprecision = “Impre”, indirectness = “Indir”

Table 5 Qualitatively summarized results, overall rating, and quality of evidence (GRADE) on *measurement error* per instrument

Instrument (reference)	Summarized result on measurement error	Overall rating	Quality of evidence	
			Quality level	Reason
EBPAS [41]	LoA < SRD on all subscales	(+)	Moderate	– 1 impre
EBP inventory [67]	MDC95 (absolute value) range from 3.5 to 7.2. Indeterminate rating (?) since MIC not defined	(?)	No grade	
EPIC [69, 70]	MDC95 (absolute value) range from 4.1 to 4.6. Indeterminate rating (?) since MIC not defined	(?)	No grade	
EBP2 [92]	SEM: ranging from 0.29 to 0.44. Indeterminate rating (?) since MIC not defined	(?)	No grade	

SEM Standard error of measurement, LoA Limits of agreement, SDCSmallest detectable change, MIC Minimal important change

Overall rating of result: (+) = sufficient result; (–) = insufficient result; (?) = indeterminate result; (±) = inconsistent results

Quality of evidence: modified GRADE approach [28, 31]. Quality levels: high, moderate, low, and very low

Reasons for downgrade: risk of bias = “RoB”, inconsistency = “Incon”, imprecision = “Impre”, indirectness = “Indir”

consistency in this review is based on the COSMIN criteria for whether or not evidence of unidimensionality on the scale exists [31]. Indeterminate (?) ratings on internal consistency alone will not lead to an instrument not being recommended in this review, since this

requires high-quality evidence of insufficient (–) measurement properties.

This review’s target population was healthcare professionals, and the number of healthcare disciplines on which an instrument was validated was one of the factors

considered when making categories of recommendations. While 17 out of the 34 included instruments were validated on two or more healthcare disciplines, 17 were validated on only one [48, 64, 65, 71, 78–80, 87, 93, 95, 96, 102, 105, 107, 109, 110]. When an instrument is validated in only one healthcare discipline, the results from a validation study may not apply if an instrument is used on a population that differs from the one on which the instrument was validated ([114], p. 230–231). Studies have shown that there may be differences between healthcare disciplines in terms of self-reported levels of EBP knowledge, attitudes, and behavior [115, 116]. It is unknown whether interdisciplinary differences in EBP knowledge, attitudes or behavior would directly affect how the items in a questionnaire are understood or to what degree they are perceived as relevant. However, knowing that a questionnaire only can be considered valid for the population on which it has been validated ([112], p.58–59), readers of this review should bear in mind that the results may not be generalizable to other populations. Readers should have a clear conception of the population on which the instrument is tested and of the population intended to target when choosing an instrument for use in future studies or clinical practice. This review's inclusion of studies from various healthcare disciplines may have contributed new knowledge to the current evidence base, identifying several valid instruments over at least two disciplines.

Most of the instruments included in this review were initially developed in English and in different English-speaking countries. Several of these instruments have been translated into other languages and used in various countries. Ideally, an instrument translation process should be conducted according to well-known guidelines to ensure that a translated instrument is valid in another language [112, 117, 118]. In this review, we did not assess the quality of the translation process, as this was not part of the COSMIN methodology recommendations used to conduct this review [26, 31]. As such, readers are advised to consider the quality of the translation process if they consider using results from studies included in this review that involved translations of instruments.

Limitations

Variations in definitions of EBP constructs between the included studies presented a challenge in the review process. Clearly defined constructs are essential to instrument development and are a prerequisite for using quantitative questionnaires to measure non-observable constructs like EBP attitudes, self-efficacy, and behavior ([112], p. 151–152). In some cases, the

differences in definitions of constructs and use of terminology made it challenging to classify the included instruments in terms of the EBP constructs measured. To meet this challenge, we classified the instruments using the CREATE framework's definitions of EBP attitudes, self-efficacy, and behavior mentioned earlier in this review [10]. For some instruments, the constructs were defined with names and terminology other than those used in the CREATE framework. The differences in definitions of constructs and use of terminology may also have affected the study selection of this review, with potentially relevant studies being overlooked and not being included. To meet this challenge, all titles and abstracts were screened by two independent review members, and a third reviewer was consulted in cases of uncertainty. Still, relevant studies and instruments may have been missed. Even though EBP theory, models, and frameworks exist, there is still a need to develop a more cohesive and clear theoretical articulation regarding EBP and the measurement of it [10, 119].

Furthermore, all the included instruments are self-reported, the most common method to measure EBP constructs. Some consider only objectively measured EBP outcomes as high-quality instruments due to the potential of recall and social desirability biases in self-reported instruments [16, 17, 22, 23]. Despite the risk of bias, others recommend using self-reported instruments as a practical option when time is an issue and an extensive, objective measurement is practically impossible [119]. In addition, it has been questioned whether the extensive focus on objectivity in EBP instruments is the only right way forward, and qualitative and mixed methods have been suggested for a richer understanding of EBP [119]. The use of a standardized and rigorous methodology (COSMIN) throughout this review may have reduced possible methodological limitations and increased the likelihood that the results and recommendations could be trusted, despite the potential risk of bias connected to self-reported instruments.

Rationale for recommendations and implications of future research

Recommendations of instruments in this review are based on the summarized results and grading of the evidence concerning the construct and population of interest. The recommendations are guided by the COSMIN methodology but are not categorized similarly [31]. The three categories are categorized based on the number of healthcare disciplines on which the instrument is validated and on the number of EBP constructs the instrument measures. Common for all three categories is that, for an instrument to be recommended, there must be

evidence of sufficient (+) content validity (any level) and no high-quality evidence of any insufficient (–) measurement properties [31]. Being recommended means that an instrument has the potential to be recommended, even though it does not have exclusively high-quality evidence of sufficient measurement properties. This aligns with research that suggests building upon existing instruments when measuring EBP attitudes, self-efficacy, and behavior [10]. Using and adapting existing instruments could also help to avoid the so-called “one-time use phenomenon,” where an instrument is developed for a specific situation and not further tested and validated in other studies ([120], p.238).

Recommendations

Instruments validated in at least two healthcare disciplines that measure two or more of the constructs in question (attitudes, behavior, self-efficacy) include the following: EBP Inventory [66], Al Zoubi questionnaire [62], EBPPAS [73], HS-EBP [85], EBP2 [89], and I-SABE [108]. Furthermore, instruments validated in at least two healthcare disciplines but that measure only one of the constructs in question include the following: EBPAS-50 [45], EBP Beliefs (single factor) [56], EBP implement (single factor) [56], EPIC [68], SE-EBP [76], and Ethiopian EBP Implement [111]. Finally, instruments validated in only one discipline that measures one or more of the constructs in question include the following: EBPQ [48], EBP (Jette) [64], EBP (Bernhardsson) [65], EBPSE [78], EBP Capability beliefs [79], HEAT [80], Quick EBP-VIK [82], ISP-D [93], EBNAQ [95], EBP Implement short [102], EIDM competence measure [107], Noor EBM [109], and EBP-CBFRI [110].

Conclusions

This review identified 34 instruments that measure healthcare professionals’ EBP attitudes, behaviors, or self-efficacy. Seventeen instruments were validated in two or more healthcare disciplines. Despite the varying quality of instrument development and content validity studies, most instruments received sufficient (+) ratings on content validity, though with a “very low” quality of evidence. The overall rating of structural validity, internal consistency, reliability, and measurement error varied, as did the quality of evidence.

Based on the summarized results, the constructs, and the population of interest, we identified several instruments that have the potential to be recommended for use in different healthcare disciplines. Future research measuring EBP attitudes, behavior, and self-efficacy should strive to build upon and further develop existing EBP instruments. In cases where new EBP instruments are

being developed, the generation of questionnaire items should include qualitative methods involving members of the target population. In addition, future research should focus on reaching a clear articulation of and a shared conception of EBP constructs.

Abbreviations

EBP	Evidence-based practice
COSMIN	Consensus-based Standards for the Selection of Health Measurement Instruments
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CREATE	Classification Rubric for Evidence-Based Practice Assessment Tools in Education
EFA	Exploratory factor analysis
CFA	Confirmatory factor analysis
CFI	Comparative fit index
RMSEA	Root mean square error of approximation
SRMR	Standardized square residual
ICC	Intraclass correlation coefficient
SEM	Standard error of measurement
LoA	Limits of agreement
SDC	Smallest detectable change
MIC	Minimal important change

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13012-023-01301-3>.

Additional file 1. Search strategy.

Additional file 2. COSMIN criteria for good measurement properties.

Additional file 3. Characteristics of the included studies and participants.

Additional file 4. Results of quality assessment and measurement properties of the individual studies.

Additional file 5. The PRISMA 2020 checklist.

Acknowledgements

The authors wish to acknowledge the following persons for their dedicated assistance with this systematic review: Thanks to Professor Astrid Bergland for collaboration and inspiration in the planning of the review. Thanks to Professor Are Hugo Pripp for reading the manuscript, and for comments regarding reporting of the measurement properties included.

Authors’ contributions

NGL, TB, and NRO developed the protocol. NGL and TB contributed to the title/abstract and full-text screening, and NRO was consulted in cases of uncertainty. NGL, TB, and NRO contributed to the quality assessment, rating, and grading of the instruments. NGL wrote the draft of the manuscript, and TB and NRO reviewed and revised the text in several rounds. All authors contributed to, reviewed, and approved the final manuscript.

Funding

Internal funding was provided by OsloMet. The funding bodies had no role in the design, data collection, data analysis, interpretation of the results, or decision to submit for publication.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Rehabilitation Science and Health Technology, Faculty of Health Science, Oslo Metropolitan University, Oslo, Norway. ²Department of Health and Functioning, Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway.

Received: 9 June 2023 Accepted: 4 September 2023

Published online: 13 September 2023

References

- 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.
- Kim SC, Ecoff L, Brown CE, Gallo AM, Stichler JF, Davidson JE. Benefits of a regional evidence-based practice fellowship program: a test of the ARCC model. *Worldviews Evid Based Nurs*. 2017;14(2):90–8.
- Melnyk BM, Gallagher-Ford L, Zellefrow C, Tucker S, Thomas B, Sinnott LT, et al. The First U S study on nurses' evidence-based practice competencies indicates major deficits that threaten healthcare quality, safety, and patient outcomes. *Worldviews Evid Based Nurs*. 2018;15(1):16–25.
- Melnyk BM, Gallagher-Ford L, Long LE, Fineout-Overholt E. The establishment of evidence-based practice competencies for practicing registered nurses and advanced practice nurses in real-world clinical settings: proficiencies to improve healthcare quality, reliability, patient outcomes, and costs. *Worldviews Evid Based Nurs*. 2014;11(1):5–15.
- Empananza JI, Cabello JB, Burls AJ. Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *J Eval Clin Pract*. 2015;21(6):1059–65.
- de Vasconcelos LP, de Oliveira Rodrigues L, Nobre MRC. Clinical guidelines and patient related outcomes: summary of evidence and recommendations. *Int J Health Governance*. 2019;24(3):230–8.
- Saunders H, Vehviläinen-Julkunen K. Key considerations for selecting instruments when evaluating healthcare professionals' evidence-based practice competencies: A discussion paper. *J Adv Nurs*. 2018;74(10):2301–11.
- da Silva TM, Costa Lda C, Garcia AN, Costa LO. What do physical therapists think about evidence-based practice? A systematic review *Man Ther*. 2015;20(3):388–401.
- 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;33(2):1–13.
- 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;11:78.
- 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):1284–303.
- 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):176–85.
- 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.
- Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, et al. Tailored interventions to address determinants of practice. *Cochrane Database Syst Rev*. 2015(4). <https://doi.org/10.1002/14651858.CD005470.pub3>.
- Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust*. 2004;180(S6):S57–60.
- 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.
- 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):91.
- Albarqouni L, Hoffmann T, Glasziou P. Evidence-based practice educational intervention studies: a systematic review of what is taught and how it is measured. *BMC Med Educ*. 2018;18(1):177.
- Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *BMJ*. 2003;326(7384):319–21.
- Hoegen PA, de Bot CMA, Echteid MA, Vermeulen H. Measuring self-efficacy and outcome expectancy in evidence-based practice: A systematic review on psychometric properties. *Int J Nurs Studies Advances*. 2021;3:100024.
- Oude Rengerink K, Zwolsman SE, Ubbink DT, Mol BW, van Dijk N, Vermeulen H. Tools to assess evidence-based practice behaviour among healthcare professionals. *Evid Based Med*. 2013;18(4):129–38.
- 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.
- 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. *Occup Ther Int*. 2016;23(2):59–90.
- 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. *J Eval Clin Pract*. 2014;20(6):767–78.
- 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;19:44.
- Terwee CB, Prinsen CAC, 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.
- Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–9.
- 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.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
- Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. [Available from: www.covidence.org].
- Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) – user manual. 2018 [Available from: <https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/>].
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. 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.
- Oude Voshaar MA, Ten Klooster PM, Glas CA, Vonkeman HE, Taal E, Krishnan E, et al. Validity and measurement precision of the PROMIS physical function item bank and a content validity-driven 20-item short form in rheumatoid arthritis compared with traditional measures. *Rheumatology (Oxford)*. 2015;54(12):2221–9.
- Elsman EBM, Mokkink LB, Langendoen-Gort M, Rutters F, Beulens J, Elders PJM, et al. Systematic review on the measurement properties of diabetes-specific patient-reported outcome measures (PROMs) for measuring physical functioning in people with type 2 diabetes. *BMJ Open Diabetes Res Care*. 2022;10(3):e002729. <https://doi.org/10.1136/bmjdc-2021-002729>.
- Schünemann H, Brożek, Guyatt, Oxman. GRADE handbook for grading quality of evidence and strength of recommendations 2013 [updated October 2013. Available from: <https://gdt.gradepro.org/app/handbook/handbook.html>].
- 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.

37. Aarons GA, McDonald EJ, Sheehan AK, Walrath-Greene CM. Confirmatory factor analysis of the Evidence-Based Practice Attitude Scale in a geographically diverse sample of community mental health providers. *Adm Policy Ment Health*. 2007;34(5):465–9.
38. Aarons GA, Glisson C, Hoagwood K, Kelleher K, Landsverk J, Cafri G. Psychometric properties and U.S. National norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychol Assess*. 2010;22(2):356–65.
39. Maessen K, van Vught A, Gerritsen DL, Lovink MH, Vermeulen H, Persoon A. Development and validation of the Dutch EBPA5-ve and EBPAQ-ve for nursing assistants and nurses with a vocational education. *Worldviews Evid Based Nurs*. 2019;16(5):371–80.
40. 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):867–76.
41. Skavberg Roaldsen K, Halvarsson A. Reliability of the Swedish version of the Evidence-Based Practice Attitude Scale assessing physiotherapist's attitudes to implementation of evidence-based practice. *PLoS ONE [Electronic Resource]*. 2019;14(11):e0225467.
42. Egeland KM, Ruud T, Ogden T, Lindstrom JC, Heiervang KS. Psychometric properties of the Norwegian version of the Evidence-Based Practice Attitude Scale (EBPAS): to measure implementation readiness. *Health research policy and systems*. 2016;14(1):47.
43. Santesson AHE, Bäckström M, Holmberg R, Perrin S, Jarbin H. Confirmatory factor analysis of the Evidence-Based Practice Attitude Scale (EBPAS) in a large and representative Swedish sample: is the use of the total scale and subscale scores justified? *BMC Med Res Methodol*. 2020;20(1):254.
44. Ashcraft RG, Foster SL, Lowery AE, Henggeler SW, Chapman JE, Rowland MD. Measuring practitioner attitudes toward evidence-based treatments: a validation study. *J Child Adolesc Subst Abuse*. 2011;20(2):166–83.
45. Aarons GA, Cafri G, Lugo L, Sawitzky A. Expanding the domains of attitudes towards evidence-based practice: the evidence based practice attitude scale-50. *Adm Policy Ment Health*. 2012;39(5):331–40.
46. Yildiz D, Fidanci BE, Acikel C, Kaygusuz N, Yildirim C. Evaluating the Properties of the Evidence-Based Practice Attitude Scale (EBPAS-50) in Nurses in Turkey. *Int J Caring Sci*. 2018;11(2):768–75.
47. Rye M, Torres EM, Friberg 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.
48. Upton D, Upton P. Development of an evidence-based practice questionnaire for nurses. *J Adv Nurs*. 2006;53(4):454–8.
49. Son Y-J, Song Y, Park S-Y, Kim J-I. A psychometric evaluation of the Korean version of the evidence-based practice questionnaire for nurses. *Contemp Nurse*. 2014;49(1):4–14.
50. Tomotaki A, Fukahori H, Sakai I, Kurokohchi K. The development and validation of the Evidence-Based Practice Questionnaire: Japanese version. *Int J Nurs Pract*. 2018;24(2):e12617.
51. Yang R, Guo JW, Beck SL, Jiang F, Tang S. Psychometric Properties of the Chinese Version of the Evidence-Based Practice Questionnaire for Nurses. *J Nurs Meas*. 2019;27(3):E117–31.
52. Zaybak A, Gunes UY, Dikmen Y, Arslan GG. Cultural validation of the Turkish version of evidence-based practice questionnaire. *Int J Caring Sci*. 2017;10(1):37–46.
53. Sese-Abad A, De Pedro-Gomez J, Bennasar-Veny M, Sastre P, Fernandez-Dominguez JC, Morales-Asencio JM. A multisample model validation of the evidence-based practice questionnaire. *Res Nurs Health*. 2014;37(5):437–46.
54. Rospodowski K, Alexandre NMC, Cornello ME. Cultural adaptation to Brazil and psychometric performance of the "Evidence-based Practice Questionnaire." *Acta Paulista De Enfermagem*. 2014;27(5):405–11.
55. Pereira RP, Guerra AC, Cardoso MJ, dos Santos AT, de Figueiredo MC, Carneiro AC. Validation of the Portuguese version of the Evidence-Based Practice Questionnaire. *Rev Lat Am Enfermagem*. 2015;23(2):345–51.
56. Melnyk BM, Fineout-Overholt E, Mays MZ. The evidence-based practice beliefs and implementation scales: psychometric properties of two new instruments. *Worldviews Evid Based Nurs*. 2008;5(4):208–16.
57. Grønvik CKU, Ødegård A, Bjørkly S. Factor analytical examination of the Evidence-Based Practice Beliefs Scale: indications of a two-factor structure. *Open J Nurs*. 2016;6(9):699–711.
58. Kerwien-Jacquier E, Verloo H, Pereira F, Peter KA. Adaptation and validation of the evidence-based practice beliefs and implementation scales into German. *Nursing Open*. 2020;7(6):1997–2008.
59. Thorsteinsson HS. Translation and validation of two evidence-based nursing practice instruments. *Int Nurs Rev*. 2012;59(2):259–65.
60. Verloo H, Desmedt M, Morin D. Adaptation and validation of the Evidence-Based Practice Belief and Implementation scales for French-speaking Swiss nurses and allied healthcare providers. *J Clin Nurs*. 2017;26(17–18):2735–43.
61. Moore JL, Friis S, Graham ID, Gundersen ET, Nordvik JE. Reported use of evidence in clinical practice: a survey of rehabilitation practices in Norway. *BMC Health Serv Res*. 2018;18(1):379.
62. 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.
63. Bernal G, Rodriguez-Soto NC. Development and psychometric properties of the evidence-based professional practice scale (EBPP-5). *P R Health Sci J*. 2010;29(4):385–90.
64. 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.
65. Bernhardsson S, Larsson ME. Measuring evidence-based practice in physical therapy: translation, adaptation, further development, validation, and reliability test of a questionnaire. *Phys Ther*. 2013;93(6):819–32.
66. Kaper NM, Swennen MH, van Wijk AJ, Kalkman CJ, van Rheenen N, van der Graaf Y, et al. The "evidence-based practice inventory": reliability and validity was demonstrated for a novel instrument to identify barriers and facilitators for Evidence Based Practice in health care. *J Clin Epidemiol*. 2015;68(11):1261–9.
67. Braun T, Ehrenbrusthoff K, Bahns C, Happe L, Kopkow C. Cross-cultural adaptation, internal consistency, test-retest reliability and feasibility of the German version of the evidence-based practice inventory. *BMC Health Serv Res*. 2019;19(1):455.
68. 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.
69. Salbach NM, Jaglal SB, Williams JL. Reliability and validity of the evidence-based practice confidence (EPIC) scale. *J Contin Educ Health Prof*. 2013;33(1):33–40.
70. Clyde JH, Brooks D, Cameron JI, Salbach NM. Validation of the Evidence-Based Practice Confidence (EPIC) Scale With Occupational Therapists. *Am J Occup Ther*. 2016;70(2):7002280010p1-9.
71. Borntrager CF, Chorpita BF, Higa-McMillan C, Weisz JR. Provider attitudes toward evidence-based practices: are the concerns with the evidence or with the manuals? *Psychiatr Serv*. 2009;60(5):677–81.
72. Park H, Ebesutani CK, Chung KM, Stanick C. Cross-cultural validation of the modified practice attitudes scale: initial factor analysis and a new factor model. *Assessment*. 2018;25(1):126–38.
73. Rubin A, Parrish DE. Development and validation of the Evidence-based Practice Process Assessment Scale: Preliminary findings. *Res Soc Work Pract*. 2010;20(6):629–40.
74. Rubin A, Parrish DE. Validation of the evidence-based practice Process Assessment Scale. *Res Soc Work Pract*. 2011;21(1):106–18.
75. Parrish DE, Rubin A. Validation of the Evidence-Based Practice Process Assessment Scale-Short Version. *Res Soc Work Pract*. 2011;21(2):200–11.
76. Chang AM, Crowe L. Validation of scales measuring self-efficacy and outcome expectancy in evidence-based practice. *Worldviews Evid Based Nurs*. 2011;8(2):106–15.
77. Oh EG, Yang YL, Sung JH, Park CG, Chang AM. Psychometric properties of Korean Version of self-efficacy of Evidence-Based practice scale. *Asian Nurs Res (Korean Soc Nurs Sci)*. 2016;10(3):207–12.
78. Tucker SJ, Olson ME, Frusti DK. Evidence-Based practice self-efficacy scale preliminary reliability and validity. *Clin Nurse Spec*. 2009;23(4):207–15.
79. Wallin L, Bostrom 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):139–48.

80. Sleutel MR, Barbosa-Leiker C, Wilson M. Psychometric testing of the health care evidence-based practice assessment tool. *J Nurs Meas*. 2015;23(3):485–98.
81. 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;14:263.
82. Paul F, Connor L, McCabe M, Ziniel S. The development and content validity testing of the Quick-EBP-VIK: A survey instrument measuring nurses' values, knowledge and implementation of evidence-based practice. *J Nurs Educ Pract*. 2016;6(5):118–26.
83. Connor L, Paul F, McCabe M, Ziniel S. Measuring nurses' value, implementation, and knowledge of evidence-based practice: further psychometric testing of the Quick-EBP-VIK Survey. *Worldviews Evid Based Nurs*. 2017;14(1):10–21.
84. Zhou C, Wang Y, Wang S, Ou J, Wu Y. Translation, cultural adaptation, validation, and reliability study of the Quick-EBP-VIK instrument: Chinese version. *J Eval Clin Pract*. 2019;25(5):856–63.
85. Fernandez-Dominguez JC, Sese-Abad A, Morales-Asencio JM, Sastre-Fullana P, Pol-Castaneda S, de Pedro-Gomez JE. Content validity of a health science evidence-based practice questionnaire (HS-EBP) with a web-based modified Delphi approach. *Int J Qual Health Care*. 2016;28(6):764–73.
86. Fernandez-Dominguez JC, de Pedro-Gomez JE, Morales-Asencio JM, Bennasar-Veny M, Sastre-Fullana P, Sese-Abad A. Health Sciences-Evidence Based Practice questionnaire (HS-EBP) for measuring trans-professional evidence-based practice: Creation, development and psychometric validation. *PLoS ONE [Electronic Resource]*. 2017;12(5):e0177172.
87. Thiel L, Ghosh Y. Determining registered nurses' readiness for evidence-based practice. *Worldviews Evid Based Nurs*. 2008;5(4):182–92.
88. Patelarou AE, Dafermos V, Broklaki H, Melas CD, Koukia E. The evidence-based practice readiness survey: a structural equation modeling approach for a Greek sample. *Int J Evid Based Healthc*. 2015;13(2):77–86.
89. 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.
90. Hu MY, Wu YN, McEvoy MP, Wang YF, Cong WL, Liu LP, et al. Development and validation of the Chinese version of the evidence-based practice profile questionnaire (EBP-Q). *BMC Med Educ*. 2020;20(1):280.
91. Panczyk M, Belowska J, Zarzeka A, Samolinski L, Zmuda-Trzebiatowska H, Gotlib J. Validation study of the Polish version of the Evidence-Based Practice Profile Questionnaire. *BMC Med Educ*. 2017;17(1):38.
92. Titlestad KB, Snibsoer AK, Stromme H, Nortvedt MW, Graverholt B, Espehaug B. Translation, cross-cultural adaptation and measurement properties of the evidence-based practice profile. *BMC Res Notes*. 2017;10(1):44.
93. Burgess AM, Chang J, Nakamura BJ, Izmirian S, Okamura KH. Evidence-based practice implementation within a theory of planned behavior framework. *J Behav Health Serv Res*. 2017;44(4):647–65.
94. Mah AC, Hill KA, Cicero DC, Nakamura BJ. A Psychometric evaluation of the intention scale for providers-direct items. *J Behav Health Serv Res*. 2020;47(2):245–63.
95. Ruzafa-Martinez M, Lopez-Iborra L, Madrigal-Torres M. Attitude towards Evidence-Based Nursing Questionnaire: development and psychometric testing in Spanish community nurses. *J Eval Clin Pract*. 2011;17(4):664–70.
96. Diermayr G, Schachner H, Eidenberger M, Lohkamp M, Salbach NM. Evidence-based practice in physical therapy in Austria: current state and factors associated with EBP engagement. *J Eval Clin Pract*. 2015;21(6):1219–34.
97. Baumann AA, Vazquez AL, Macchione AC, Lima A, Coelho AF, Juras M, et al. Translation and validation of the evidence-based practice attitude scale (EBPAS-15) to Brazilian Portuguese: Examining providers' perspective about evidence-based parent intervention. *Child Youth Serv Rev*. 2022;136. <https://doi.org/10.1016/j.chilcyouth.2022.106421>.
98. Ayhan Baser D, Agadayi E, Gonderen Cakmak S, Kahveci R. Adaptation of the evidence-based practices attitude scale-15 in Turkish family medicine residents. *Int J Clin Pract*. 2021;75(8):e14354.
99. Van Giang N, Lin SY, Thai DH. A psychometric evaluation of the Vietnamese version of the Evidence-Based Practice Attitudes and Beliefs Scales. *Int J Nurs Pract*. 2021;27(6):e12896.
100. Szota K, Thielemann JFB, Christiansen H, Rye M, Aarons GA, Barke A. Cross-cultural adaptation and psychometric investigation of the German version of the Evidence Based Practice Attitude Scale (EBPAS-36D). *Health Res Policy Systems*. 2021;19(1):90.
101. Fajarini M, Rahayu S, Setiawan A. The Indonesia version of evidence-based practice questionnaire (EBPQ): Translation and Reliability. *Indones Contemp Nurs J*. 2021;5(2):42–8.
102. Melnyk BM, Hsieh AP, Gallagher-Ford L, Thomas B, Guo J, Tan A, et al. Psychometric Properties of the Short Versions of the EBP Beliefs Scale, the EBP Implementation Scale, and the EBP Organizational Culture and Readiness Scale. *Worldviews on Evidence-Based Nursing*. 2021;18(4):243–50.
103. Ferreira RM, Ferreira PL, Cavalheiro L, Duarte JA, Gonçalves RS. Evidence-based practice questionnaire for physical therapists: Portuguese translation, adaptation, validity, and reliability. *J Evidence-Based Healthcare*. 2019;1(2):83–98.
104. Belowska J, Panczyk M, Zarzeka A, Iwanow L, Cieslak I, Gotlib J. Promoting evidence-based practice - perceived knowledge, behaviours and attitudes of Polish nurses: a cross-sectional validation study. *Int J Occup Saf Ergon*. 2020;26(2):397–405.
105. Ruzafa-Martinez M, Fern-Salazar S, Leal-Costa C, Ramos-Morcillo AJ. Questionnaire to Evaluate the Competency in Evidence-Based Practice of Registered Nurses (EBP-COQ Prof®): Development and Psychometric Validation. *Worldviews Evidence-Based Nursing*. 2020;17(5):366–75.
106. Schetaki S, Patelarou E, Giakoumidakis K, Trivli A, Kleisiaris C, Patelarou A. Translation and Validation of the Greek Version of the Evidence-Based Practice Competency Questionnaire for Registered Nurses (EBP-COQ Prof®). *Nursing Reports*. 2022;12(4):693–707.
107. Belita E, Yost J, Squires JE, Ganann R, Dobbins M. Development and content validation of a measure to assess evidence-informed decision-making competence in public health nursing. *PLoS ONE [Electronic Resource]*. 2021;16(3):e0248330.
108. Ruano ASM, Motter FR, Lopes LC. Design and validity of an instrument to assess healthcare professionals' perceptions, behaviour, self-efficacy and attitudes towards evidence-based health practice: I-SABE. *BMJ Open*. 2022;12. <https://doi.org/10.1136/bmjopen-2021-052767>.
109. Norhayati MN, Nawi ZM. Validity and reliability of the Noor Evidence-Based Medicine Questionnaire: A cross-sectional study. *PLoS ONE [Electronic Resource]*. 2021;16(4):e0249660.
110. Abuadas MH, Albikawi ZF, Abuadas F. Development and Validation of Questionnaire Measuring Registered Nurses' Competencies, Beliefs, Facilitators, Barriers, and Implementation of Evidence-Based Practice (EBP-CBFRI). *J Nurs Meas*. 2021;13:13.
111. Dessie G, Jara D, Alem G, Mulugeta H, Zewdu T, Wagnaw F, et al. Evidence-based practice and associated factors among health care providers working in public hospitals in Northwest Ethiopia During 2017. *Curr Ther Res Clin Exp*. 2020;93:100613.
112. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press; 2011. <https://doi.org/10.1017/CBO9780511996214>.
113. McNeish D. Thanks coefficient alpha, we'll take it from here. *Psychol Methods*. 2018;23(3):412–33.
114. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. New York, New York: Oxford University Press; 2015.
115. McEvoy MP, Williams MT, Olds TS. Evidence based practice profiles: differences among allied health professions. *BMC Med Educ*. 2010;10:69.
116. Upton D, Upton P. Knowledge and use of evidence-based practice by allied health and health science professionals in the United Kingdom. *J Allied Health*. 2006;35(3):127–33.
117. Mokkink LB, Prinsen CA, Patrick D, Alonso J, Bouter LM, Vet HCD, et al. *Cosmin Study design checklist for patient-reported outcome measurement instruments [PDF]*. <https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/> 2019 [Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-design-checklist_final.pdf#].

118. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–91.
119. Roberge-Dao J, Maggio LA, Zaccagnini M, Rochette A, Shikako K, Boruff J, et al. Challenges and future directions in the measurement of evidence-based practice: Qualitative analysis of umbrella review findings. *J Eval Clin Pract*. 2023;29:218–27.
120. Brownson RC, Colditz GA, Proctor EK. Dissemination and implementation research in health : translating science to practice. New York, NY: Oxford University Press; 2017.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

