Article:
Hong, Q.N., Pluye, P., Fàbregues, S. et al. (10 more authors) (2019) Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. Journal of Clinical Epidemiology. ISSN 0895-4356

https://doi.org/10.1016/j.jclinepi.2019.03.008

ELSEVIER

Journal of Clinical Epidemiology xx (2019) 1−11

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study

Quan Nha Hong[a], Pierre Pluye[a,*], Sergi Fàbregues[b], Gillian Bartlett[a], Felicity Boardman[c], Margaret Cargo[d], Pierre Dagenais[e], Marie-Pierre Gagnon[f], Frances Griffiths[c], Belinda Nicolau[g], Alicia O'Cathain[h], Marie-Claude Rousseau[i], Isabelle Vedel[a]

[a]*Department of Family Medicine, McGill University, 5858 Chemin de la Côte-des-Neiges, Suite 300, Montréal, Québec, H3S 1Z1, Canada*
[b]*Department of Psychology and Education, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018, Barcelona, Spain*
[c]*Warwick Medical School − Division of Health Sciences, University of Warwick, Coventry, CV4 7AL, England*
[d]*Health Research Institute, University of Canberra, Canberra, ACT, 2601, Australia*
[e]*Faculté de médecine et des sciences de la santé, Université de Sherbrooke, 3001, 12e Avenue Nord, Sherbrooke, Québec, J1H 5N4, Canada*
[f]*Faculté des sciences infirmières, Université Laval, 1050, avenue de la Médecine, Québec, Québec, G1V 0A6, Canada*
[g]*Faculty of Dentistry, Division of Oral Health and Society Research, McGill University, 2001 McGill College, suite 500, Montréal, Québec, H3A 1G1, Canada*
[h]*Medical Care Research Unit, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, S1 4DA, UK*
[i]*INRS−Institut Armand-Frappier Research Centre, 531, boulevard des Prairies, Laval, Québec, H7V 1B7, Canada*

Accepted 6 March 2019; Published online xxxx

Abstract

Objective: The mixed methods appraisal tool (MMAT) was developed for critically appraising different study designs. This study aimed to improve the content validity of three of the five categories of studies in the MMAT by identifying relevant methodological criteria for appraising the quality of qualitative, survey, and mixed methods studies.

Study Design and Setting: First, we performed a literature review to identify critical appraisal tools and extract methodological criteria. Second, we conducted a two-round modified e-Delphi technique. We asked three method-specific panels of experts to rate the relevance of each criterion on a five-point Likert scale.

Results: A total of 383 criteria were extracted from 18 critical appraisal tools and a literature review on the quality of mixed methods studies, and 60 were retained. In the first and second rounds of the e-Delphi, 73 and 56 experts participated, respectively. Consensus was reached for six qualitative criteria, eight survey criteria, and seven mixed methods criteria. These results led to modifications of eight of the 11 MMAT (version 2011) criteria. Specifically, we reformulated two criteria, replaced four, and removed two. Moreover, we added six new criteria.

Conclusion: Results of this study led to improve the content validity of this tool, revise it, and propose a new version (MMAT version 2018).

Keywords: Quality appraisal; Delphi technique; Systematic review; Qualitative research; Surveys; Mixed methods research

1. Introduction

Systematic reviews are considered among the best available sources of research evidence and are increasingly relied on to inform decision-making [1]. The past 40 years have seen increasingly rapid methodological advances in the field of systematic reviews and research synthesis. Initial developments mainly focused on meta-analysis for addressing questions on the effectiveness of interventions, and the emphasis was on randomized controlled trials [2,3]. Since the early 2000s, researchers have shown a growing interest in systematic mixed studies reviews,

Conflict of interest statement: Quan Nha Hong, OT, MSc, PhD. This manuscript was written while she was a PhD candidate and held a Doctoral Fellowship Award from the Canadian Institutes of Health Research (CIHR). Pierre Pluye, MD, PhD, Full Professor, holds a Senior Investigator Award from the Fonds de recherche du Québec−Santé (FRQS) and is the Director of the Methodological Development Platform of the Quebec-SPOR SUPPORT Unit, which is funded by the CIHR, the FRQS, and the Quebec Ministry of Health.

* Corresponding author. Department of Family Medicine, McGill University, 5858 Chemin de la Côte-des-Neiges, Suite 300, Montréal, Québec, Canada H3S 1Z1. Tel.: +1-514-398-8483; fax: +1-514-398-4202.

E-mail address: pierre.pluye@mcgill.ca (P. Pluye).

https://doi.org/10.1016/j.jclinepi.2019.03.008

What is new?

Key findings

- A revised version of the Mixed Methods Appraisal Tool (MMAT) was developed and includes 25 criteria on five categories of studies. This critical appraisal tool was developed to assess the methodological quality of various study designs, including mixed methods studies.

What this adds to what was known

- The first version of the MMAT was published in 2009. We further developed the MMAT based on experts' opinions to improve its content validity.
- This study adds to the literature on the quality of qualitative, survey and mixed methods research, which is still sparse and lacking consensus.

What is the implication and what should change now?

- A new content validated version of the MMAT was developed and can be useful for the critical appraisal process in systematic reviews combining qualitative and quantitative evidence.

which combine quantitative, qualitative, and mixed methods studies to address other types of review questions concerned with, for instance, the acceptability of an intervention, participants' satisfaction, or barriers to implementation (see Supplementary File 1). Systematic mixed studies reviews are particularly useful for providing in-depth answers to complex clinical problems and practical concerns. Several challenges, however, are encountered in these reviews because of the heterogeneity of included study designs. One of these challenges pertains to the critical appraisal of included studies.

Critical appraisal consists in a systematic and careful examination of studies to ensure they are trustworthy, valid, and reliable [4,5]. It is an essential step in systematic reviews to ensure that their recommendations and conclusions reflect the quality of the evidence reviewed [6]. Since reviewers' judgment of a same study can vary greatly, critical appraisal tools have been developed to help reviewers appraise study quality in a more consistent, transparent, and reproducible way [7−9]. A critical appraisal tool (also named quality assessment tool or risk of bias tool) is a scale or checklist in which a list of criteria/domains is suggested to appraise the quality of a study. Extant reviews of critical appraisal tools have identified over 500 tools (see Supplementary File 1). Most of these tools are specific to a particular research design or method. It is, thus, complex and time consuming to conduct systematic mixed studies reviews as reviewers must search for and learn how to use several different tools to complete the critical appraisal of the qualitative, quantitative, and/or mixed methods studies included in each review.

To address the challenge of critical appraisal in systematic mixed studies reviews, a unique tool for assessing the quality of different study designs was developed: the Mixed Methods Appraisal Tool (MMAT) [10]. The MMAT was first published in 2009 and has five sets of criteria for: (a) qualitative (such as case study and grounded theory), (b) randomized controlled trials, (c) nonrandomized (such as cohort studies and case-control studies), (d) quantitative descriptive (such as surveys and case series), and (e) mixed methods studies. When appraising mixed methods studies, three sets of criteria are assessed in no particular order: (a) the qualitative set, (b) a quantitative set (either randomized controlled, nonrandomized or quantitative descriptive studies), and (c) the mixed methods set. In doing so, the MMAT acknowledges the methodological distinctive characteristics specific to each component used in mixed methods studies (i.e., qualitative, quantitative, and mixed methods) [11].

Previous studies on the interrater reliability of the MMAT reported that agreement scores ranged from poor to perfect [12,13]. This suggests the need for clarification of some criteria in the MMAT, particularly those related to qualitative and nonrandomized studies, for which lower agreement was observed. In addition, in interviews conducted with MMAT users to explore their views and experiences of the MMAT, concerns were raised about whether the tool included enough criteria to judge the quality of studies and criteria that were difficult to judge, in particular the criteria for qualitative and mixed methods studies [14]. This suggests a need to improve the content validity of the MMAT. The content validity of an assessment tool is defined as the degree to which criteria are relevant to and representative of their targeted construct [15]. A conceptual framework on the quality appraisal in systematic mixed studies reviews was developed in which three dimensions of quality were presented: reporting, conceptual, and methodological [16]. Reporting quality relates to the transparency, accuracy, and completeness of the information provided in a paper. Conceptual quality concerns the insight that can be gained about the phenomenon of interest. The methodological quality concerns the validity or trustworthiness of a study and is related to the methodology and methods used and how biases were minimized. In the MMAT, the targeted construct is the methodological quality of studies appraised in systematic mixed studies reviews.

Currently, the existing literature on critical appraisal has focused, for the most part, on randomized controlled trials, cohort studies, and/or case-control studies, and several validated tools can be found for these study designs. This literature will inform the criteria on randomized controlled trials and nonrandomized studies to revise in the MMAT. However, for other designs, such as qualitative, survey, and mixed

methods, critical appraisal is more challenging because validated tools are rare and there is no clear consensus on how their quality assessment should be performed [17−19].

The objective of this study was to improve the content validity of the MMAT by identifying the most relevant methodological criteria for appraising the quality of qualitative, survey, and mixed methods studies. This study focused on these three categories of studies because of the scarcity of literature and lack of consensus.

2. Methods

Two phases were conducted: (a) a literature review to identify existing criteria and (b) a modified e-Delphi technique. The Delphi technique is used to reach consensus among a group of experts [20] and is particularly suitable to build consensus on issues that have limited or contradictory evidence [21]. It has been used for the development of other critical appraisal tools for different types of studies such as prognostic studies, case series studies, cross-sectional studies, studies on measurement properties, and randomized controlled trials (see Supplementary File 1). The Delphi technique is characterized by two or more rounds of questionnaires with controlled feedback, statistical group response, and anonymity [20]. There are different types of Delphi designs [20]. We used a modified e-Delphi, meaning that the Delphi was administered via an online web survey and used preselected methodological criteria in the first round.

2.1. Phase 1: literature review

To identify methodological criteria, we performed a literature review of critical appraisal tools for qualitative, surveys, and mixed methods studies. In the MMAT, because surveys are part of the quantitative descriptive studies category, we also included tools that were related to cross-sectional and prevalence studies.

2.1.1. Sources

Two main literature sources were used. The first was a review of systematic mixed studies reviews that was carried out in 2015 [22]. In this review, six databases (MEDLINE, PsycINFO, Embase, CINAHL, AMED, and Web of Science) were searched from inception of each database until December 8, 2014 and analyzed 459 reviews. The second was 15 reviews on critical appraisal tools identified from citation tracking of tools found in the first source and from reviews known to the authors of this paper (see Supplementary File 1). Also, based on the findings of our review of systematic mixed studies reviews [22], we also considered tools often used and which were developed by three leading international institutions: Critical Appraisal Skills Programme, Joanna Briggs Institute, and National Institute for Health and Clinical Excellence.

2.1.2. Selection criteria

Critical appraisal tools assessing methodological quality were retained, whereas tools limited to the quality of reporting of studies were excluded. Tools that included both reporting and methodological quality criteria were retained and only the methodological quality criteria were considered. We only retained appraisal tools that provided a clear description of their development with a group of experts or that had been subject to validity or reliability testing.

2.1.3. Identification of items

For each retained appraisal tool, all the criteria were extracted and entered in a spreadsheet by one person (QNH). Two team members (QNH, PP) independently screened the list to include methodological quality criteria. The following were excluded: criteria limited to the quality of reporting (e.g., the response rate is reported); generic criteria, that is, criteria that were related to the general steps for conducting any research study (e.g., the problem is accurately depicted or ethical issues are adequately considered); and criteria that were specific to a topic (e.g., the ethnic composition of the population studied is recorded or the topic is relevant to primary health care). Duplicates and criteria on the same concept were removed (e.g., reflexivity of the account and evidence of reflexiveness in the process). The preliminary list was sent to all members of the research team (authors of this paper) who had backgrounds in qualitative, epidemiology, and mixed methods studies. They were asked to review the list, identify the criteria that were unclear, and suggest modifications, if necessary. They were also asked to suggest criteria they felt were missing from the list.

2.2. Phase 2: two-round modified e-Delphi study

Three method-specific panels of experts were asked to complete two rounds of Delphi questionnaires to identify the most relevant methodological criteria for critical appraisal. Relevance was defined as the appropriateness of the elements to the targeted construct [15]. In this study, the targeted construct was the methodological quality of studies.

2.2.1. Sample

For each panel, a purposeful sample of international experts was constituted. An expert is defined as an individual with knowledge and skills in a specific area [23]. For the purposes of this e-Delphi, the experts were researchers working in an academic or research institution with research interests in the methodological development of either qualitative, survey, or mixed methods studies. To identify the experts, the lead author performed a search of books and methodological papers in Google Scholar, the McGill Library catalog, and Amazon. Then, the biographies of publications' authors were consulted on the World Wide Web to verify their research design expertise (e.g., by

checking their research interest and expertise, courses taught, and scientific publications). The lead author compiled the list of experts, categorized by research design, and submitted it to the full research team, asking members to add any missing experts. A total of 196 experts (i.e., potential participants) were retained.

2.2.2. Data collection and analysis

The questionnaires were put online using the LimeSurvey software hosted on the McGill University server. Pilot testing of the online questionnaires was conducted with one professor, two graduate students, and one research associate to obtain feedback regarding the clarity of the instructions, ease of completing the questionnaires, technical difficulties encountered, and to estimate the time needed to complete the task.

In Round-one, the experts were asked to rate the relevance of each criterion. A 5-point Likert scale was used, ranging from 1 = not at all relevant to 5 = extremely relevant. Space was included at the end of the questionnaire for participants to provide comments and suggestions. A 1-month turnaround time was given for panel members to complete the questionnaire. Based on the comments provided in Round-one, some criteria were modified and new criteria were added. A summary table of the results including group ratings and comments obtained in this round was prepared. This table was used to provide controlled feedback and statistical group response to participants, two important characteristics of the Delphi technique [24].

For Round-two, each participant was sent the summary table including a reminder of their responses and a new questionnaire to complete. The participants were asked to (re)rate all criteria using the same 5-point Likert scale. In addition, a ''cannot answer'' response category was added (at the request of participants). Space was provided at the end of each question for comments and suggestions. The data of Round-two were summarized by calculating an agreement index. For each item, the number of experts rating criteria as very relevant or extremely relevant was divided by the total number of experts. For each item, we considered that consensus had been reached if the agreement index was 0.80 or more.

We used the agreement indexes and the comments from Round-two as well as the literature review on critical appraisal tools to inform the revision of the MMAT. Specifically, we verified if the criteria in the current version of the MMAT (version 2011) were among those with an agreement score $\geq$ 0.80. If not, we considered how they could be modified or replaced with new ones on similar concepts. Experts' comments were used to reformulate some criteria.

2.3. Ethics statement

This project was approved by the Institutional Review Board of the Faculty of Medicine Research and Graduate Studies Offices from McGill University (ethics certificate number # A05-E26-15B). An electronic consent form was included in the questionnaire of Round-one. All experts provided informed consent to participate in this study and to be acknowledged in this paper. The responses were kept anonymous to the panel, and no personally identifiable information was presented in the data file used for the analysis.

3. Results

3.1. Phase 1: literature review

A total of 18 critical appraisal tools were retained (see Supplementary File 1): nine for qualitative studies, seven for surveys, including cross-sectional and prevalence studies, and two that included criteria for judging the quality of qualitative and quantitative studies. Because only one tool with criteria specific to mixed methods studies was retained [10], the results of a recent literature review performed by a member of our research team on the quality of mixed methods studies were used [25]. In this latter review, the authors analyzed 64 articles on the quality of mixed methods studies and identified 46 criteria [25].

Overall, 383 criteria were extracted from the included literature (238 for qualitative studies, 99 for surveys, and 46 for mixed methods studies), of which 286 (75%) were removed because they were either duplicate, generic, topic related, or limited to reporting quality. The remaining 97 criteria were presented to the research team to assess their comprehensiveness and clarity; 38 were removed because they were not clear or similar to other criteria. Also, a member of the research team suggested adding one criterion on the content validity for surveys. The 60 retained criteria included 20 for qualitative studies, 20 for quantitative descriptive studies, and 20 for mixed methods studies.

3.2. Phase 2: modified e-Delphi

Table 1 presents the number of participants in each round of the modified e-Delphi and for each of the three panels. A total of 73 experts from 11 different countries participated in Round-one: Australia (n = 2), Belgium (n = 3), Canada (n = 11), England (n = 9), Estonia (n = 1), Germany (n = 1), the Netherlands (n = 4), Norway (n = 1), Spain (n = 1), Switzerland (n = 1), and the United States of America (n = 39).

Table 1. Number of experts in each round of the e-Delphi study

Panel	Invitation	Round-one	Round-two
Qualitative	72	26	21
Survey	66	21	15
Mixed methods	58	26	20
Total	196	73	56

Based on the results of Round-one, of the initial 60 criteria, six criteria were removed, 25 criteria were reformulated, and eight new criteria were added; two qualitative studies criteria were removed, 15 modified, and two added; three survey criteria were removed, nine modified, and three added; one mixed methods studies criterion was removed, one modified, and three added. Thus, the Round-two questionnaires included 62 criteria: 21 criteria for qualitative studies, 20 criteria for surveys, and 21 criteria for mixed methods studies. The new questionnaires were sent to the 73 participants from Round-one, 56 of whom completed Round-two (Table 1). Consensus was reached for six qualitative studies criteria, eight survey criteria, and seven mixed methods studies criteria. The results of Round-two are presented in Tables 2−4.

3.3. Update of the mixed methods appraisal tool (MMAT)

In light of the results, 8 of the 11 criteria in the MMAT (version 2011) were modified: two were reformulated, four replaced, and two removed. Moreover, six new criteria were added. Supplementary File 2 presents the initial and new criteria.

3.3.1. Qualitative studies criteria

Two criteria included in the MMAT (version 2011) were not considered among the most relevant criteria to appraise in this modified e-Delphi: criterion 1.3 on the influence of the context and criterion 1.4 about of researchers' reflexivity. Some experts considered that this latter criterion might not always be reported, given space limitations in journal publications. Inadequate reporting in qualitative studies is an important barrier to critical appraisal [26]. Based on these results, the research team decided to replace criteria 1.3 and 1.4 by three new criteria that reached high level of consensus in Round-two: one on the relevance of the qualitative approach to address the research question (Table 2, criterion #1); one on the coherence between data sources, collection, analysis, and interpretation (Table 2, criterion #14); and one on the interpretation of results (Table 2, criterion #19).

Two criteria concerning the interpretation of results achieved a high level of consensus (Table 2, criteria #18

Table 2. Delphi results with experts in qualitative studies (n = 21)

Criteria[a]	Agreement index
1. A qualitative approach is appropriate to answer the research question.	**1.00**
2. The methods were adapted to fit the context of the study.	0.71
3. The role(s) of researcher(s) are discussed in terms of their assumptions and position as insider/outsider relative to the phenomenon, participants, and/or setting.	0.67
4. The researcher's involvement in the data collection and analysis is appropriate for the method used.	0.57
5. The sampling strategy is appropriately justified.	0.71
6. The sample size is appropriate for the research design.	0.43
7. The sample represents the diversity of the people for whom the study is relevant.	0.24
8. The characteristics of the sample relevant to the interpretation of the findings are appropriately described.	0.71
9. The sites of recruitment are appropriate for addressing the purpose of the study.	0.38
10. The sources of qualitative data (such as archives, documents, participant observation, etc.) are appropriate to address the research question.	**0.86**
11. The qualitative data collection methods are most appropriate to address the research question.	0.76
12. The qualitative data analysis methods are appropriately addressed.	0.67
13. Appropriate explanation is given for how findings (such as themes, concepts, categories, etc.) were derived from the data.	**0.81**
14. There is coherence between qualitative data sources, collection, analysis, and interpretation.	**0.81**
15. Strategies (such as prolonged engagement, peer review, etc.) are used to strengthen the findings.	0.71
16. Appropriate consideration is given to how findings relate to the context (such as the setting where the data were collected, etc.).	0.71
17. The influence of the researcher(s) on the data collection and analysis, results and interpretation is appropriately considered.	0.76
18. The interpretation of results is plausible.	**0.86**
19. The interpretation of results is sufficiently substantiated with data.	**0.90**
20. Any relevant epistemological or theoretical framework used is appropriately explained and justified.	0.62
21. The contextual relations between the researcher(s) and the participants (and/or materials) of research are appropriately addressed.	0.43

[a] Criteria in bold had an agreement index ≥0.80.

Table 3. Delphi results with experts in survey studies (n = 15)

Criteria[a]	Agreement index
1. The target population is clearly defined.	**1.00**
2. The study participants and the setting are described in detail.	**1.00**
3. The list from which the sample is drawn is appropriate for answering the research question.	**1.00**
4. The sampling strategy is relevant to address the research question.	**0.87**
5. The sample is representative of the target population for the main relevant variables.	**0.87**
6. The sample size is appropriate considering the population under study (such as population size, expected response rate, etc.).	0.53
7. The sample size is based on prestudy considerations of statistical power.	0.40
8. The same methods of data collection are used for all participants.	0.13
9. Standard instruments are used for the measurement of the variables.	0.33
10. The choice of variables is based on their content validity.	0.73
11. The survey instrument was pretested.	0.60
12. The survey instrument is reliable.	0.66
13. The survey instrument is valid.	0.66
14. The statistical analysis is appropriate to answer the research question.	**1.00**
15. The sampling bias is adequately addressed in the analysis.	**0.87**
16. Confounding factors are identified and accounted for in the analysis.	**0.80**
17. The response rate is acceptable (60% or above).	0.47
18. There is no significant difference in relevant sociodemographic characteristics between the respondents and the nonrespondents.	0.40
19. Weighting for nonresponse is carried out.	0.60
20. A clear justification for using survey method is provided.	0.46

[a] Criteria in bold had an agreement index ≥0.80.

and 19). In Round-one, they were combined, but experts requested they be separated because they address two different constructs (plausibility of finding vs. sufficient substantiation of findings). The latter criterion was retained for the new version of the MMAT because the agreement index was slightly higher than the former and plausibility might be more difficult to judge.

In addition, modifications were made to the first two qualitative criteria. The word ‘‘interviews’’ was added to criterion 1.1, and the word ‘‘relevant’’ was replaced by ‘‘adequate’’. Criterion 1.2 on analysis was reformulated and the word ‘‘objective’’ was removed (see Supplementary File 2).

3.3.2. Survey criteria

Experts reached consensus on eight criteria. Some of these criteria addressed similar constructs and were thus combined. For example, to judge if a sample is representative of the target population (Supplementary File 2, criterion 4.2 in the MMAT), the target population needs to be clearly defined (Table 3, criterion #1), and the study participants and setting need to be detailed (Table 3, criterion #2).

Concerning measurement bias, we included six criteria in the questionnaire but none achieved consensus. Several experts mentioned that the criteria on measurement could be useful in some circumstances but not all. In the literature, measurement error is an important aspect to consider when conducting a survey [27]. Thus, no change was made to criterion 4.3 in the MMAT.

The original MMAT criterion on response rate was replaced with one on nonresponse bias (Supplementary File 2, criterion 4.4). The appropriateness of the response rate for surveys is often requested in appraisal tools. Some will use a cutoff (e.g., 60%). However, the experts mentioned that the cutoff value is arbitrary and that less emphasis should be put on a norm. Instead the focus should be placed on nonresponse bias. This concurs with studies reporting a weak association between response rate and nonresponse bias [28].

One criterion on the appropriateness of statistical analysis reached consensus for relevance by the experts (Table 3, criterion #14) and was added to the MMAT. Also, criterion #16 on confounding factors being accounted for in the analysis achieved consensus among the experts. This criterion was not added in the section quantitative descriptive studies of the MMAT because it is mainly applicable for analytical surveys. Analytical studies are addressed in another section of the MMAT.

3.3.3. Mixed methods studies criteria

All three MMAT criteria pertaining to mixed methods were replaced. The first criterion on the relevance of

Table 4. Delphi results with experts in mixed methods studies (n = 20)

Criteria[a]	Agreement index
1. A mixed methods research question (or purpose statement) is formulated.	0.60
2. A clear rationale is provided for using a mixed methods design to address the research problem and questions.	**0.95**
3. Key literature on mixed methods is reviewed in support of the mixed methods approach chosen by the authors.	0.20
4. The mixed methods design is consistent with the epistemological assumptions of the study.	0.30
5. Methods were selected to minimize shared bias.	0.25
6. Quantitative and qualitative components of the study are effectively integrated.	**0.85**
7. The type of integration of the quantitative and qualitative components matches the mixed methods design	0.70
8. The epistemological, ontological, and teleological stances of the researcher that underlie the quantitative and quantitative approaches are successfully combined	0.10
9. Strategies for integrating phases, results, and/or data are adequately performed.	**0.90**
10. Methods are implemented in a way that remains true to the mixed methods design.	0.70
11. The qualitative and quantitative components are linked in a cohesive and logical manner.	**0.85**
12. Divergences and inconsistencies between quantitative and qualitative results are adequately addressed.	**0.90**
13. Inferences derived from the quantitative and qualitative results are adequately incorporated in the meta-inferences regarding the entire study.	**0.90**
14. Meta-inferences regarding the entire study are consistent with the rationale given for using a mixed methods design.	0.50
15. The study contributes to advancing the field of mixed methods research.	0.10
16. The added value gained from using a mixed methods design in this study is described.	0.50
17. The strengths and weaknesses of methods optimize the breadth and depth of the study.	0.30
18. Threats to the trustworthiness of quantitative, qualitative, and mixed methods are identified and adequately addressed.	**0.80**
19. Rigorous procedures for data collection and analysis are used in quantitative and qualitative components.	0.75
20. The study purposefully seeks out diverse perspectives (interpretive comprehension).	0.35
21. The mixed methods study generated findings and insights that would not have been possible with a mono-method study.	0.55

[a] Criteria in bold had an agreement index ≥0.80.

research design (Supplementary File 2, criterion 5.1 in the MMAT) was replaced with a criterion on rationale (Table 4, criterion #2).

The second criterion on integration (Supplementary File 2, criterion 5.2 in the MMAT) was reformulated. Several items on integration reached consensus (Table 4, criteria #6, 9, 11). For the MMAT, we retained the criterion #6 (quantitative and qualitative components of the study are effectively integrated) because it was also mentioned in other studies as among the most prevalent criterion for assessing the quality of mixed methods studies [17,29]. Also, some experts suggested avoiding the reference to qualitative and quantitative components in the formulation of the criteria. We replaced ''quantitative and qualitative components'' by ''different components''. In mixed methods studies, integration can be considered at different levels (e.g., philosophical, methodology, methods, data collection, and analysis techniques), and one expert suggested being more precise on what is being integrated. In a review on mixed methods studies, Pluye et al. [30] identified nine strategies for integrating phases, results, or data. Also, Fetters, Curry, and Creswell [31] identified three integration levels (design, methods, and interpretation/reporting). Because integration can vary depending on how the study was conducted, no further information was added in the criterion to keep it comprehensive.

The third criterion on limitations in mixed methods studies (Supplementary File 2, criterion 5.3) was replaced with one on meta-inferences (Table 4, criterion #13) and one on divergences (Table 4, criterion #12). Several experts mentioned that the term ''meta-inference'' was unclear. This criterion was reformulated as follows: The outputs of the integration of qualitative and quantitative components are adequately interpreted (Supplementary File 2).

One criterion was added about the trustworthiness of the qualitative and quantitative components (Table 4, criterion #18). Yet, the use of the term ''trustworthiness'' did not reach consensus among the experts (some considered this term to be associated with qualitative research). Other terms were suggested such as legitimation, validity credibility, and integrity. To avoid entering into a semantic debate, we decided to reformulate this criterion based on the work of Fàbregues, Paré, and Meneses [29]: The different components adhere to the quality criteria of each tradition of the

methods involved. As mentioned earlier, the MMAT was conceived as a building block. Thus, the appraisal of the quality of each component in mixed methods studies is done using the criteria from the other sets in the MMAT.

4. Discussion

A framework for developing assessment tools has been proposed in which three main stages are defined: initial steps, tool development, and dissemination [32]. This study is situated in the tool development stage by generating and seeking for consensus on criteria for three of the five study categories included in the MMAT (qualitative, survey, and mixed methods studies). We used a modified e-Delphi technique to identify the most relevant criteria for appraising the quality of these three categories. Consensus was reached for six criteria related to qualitative studies, eight for surveys, and seven for mixed methods studies. Results of this study improved the content validity of the MMAT, informed its revision, and led to propose a new version (MMAT version 2018).

Three main changes have been made to the MMAT. In the previous version, the MMAT had four criteria for each category of studies. Based on our results, the revised version is composed of five criteria for each category of studies, and changes were made in some criteria of the MMAT (see Supplementary File 2). Another change concerns the overall numerical score. In the previous version, an overall score could be calculated by counting the number of criteria rated ''yes''. Currently, in the literature on critical appraisal tools, it is discouraged to calculate an overall score because it does not provide information on what aspects of studies are problematic and provide equal weight to all criteria [33−37]. On this basis, it was decided to remove the overall numerical score from the MMAT. Instead, it is advised to provide a detailed presentation of the ratings of the criteria to better inform the quality of the included studies and encourage performing sensitivity analysis. Third, changes were made in the user manual and an algorithm was added to help MMAT users choose the set(s) of criteria to use. The algorithm was developed based on existing algorithms of quantitative study designs (see Supplementary File 1). The version 2018 of the MMAT is available at this website: http://mixedmethodsappraisaltoolpublic.pbworks.com/ (see Appendix 1).

The results of the critical appraisal of individual studies can be used to assess the overall quality of evidence and strength of the recommendations, that is, to judge how much confidence to place in the body of evidence. Several approaches for rating the overall quality of evidence have been developed, such as Grading of Recommendations, Assessment, Development and Evaluations (GRADE) [38] and GRADE-Confidence in the Evidence from Reviews of Qualitative research (CERQual) [39]. In these approaches, the methodological quality of individual studies (or risk of bias) is one factor that is considered among others such as the relevance of the evidence to answer the review question (indirectness), variation across studies (inconsistency), and random error on evidence (imprecision).

There is a need to further content validate the criteria identified in this study, particularly for surveys. In this study, no criteria related to measurement and response rate biases in surveys made consensus (Table 3). This might be due to the fact that diverse sources can influence measurement errors (e.g., questionnaire, data collection method, interviewer, and respondent) [27] and can vary from one study to the other. As for response rate, different indicators can be used to judge nonresponse bias such as identifying the reasons for nonresponse, determining if the respondents and nonrespondents differ on the survey variable of interest and weighting for nonresponse [27]. Although no specific criteria on measurement and response rate reached high level of consensus, the research team decided not to exclude these two biases from the MMAT because they are often mentioned in the literature [27,40,41]. Further content validation work is needed to refine these criteria. Also, in the MMAT version 2011, surveys are included in the broad ''quantitative descriptive studies'' category. We focused on surveys because they are often included in systematic mixed studies reviews, the existing tools have not been developed with experts, and surveys are among the most commonly used methods in mixed methods studies [42]. Subsequent research should verify if the new criteria are applicable to other quantitative descriptive study designs.

Developing clear critical appraisal criteria is challenging. Experts provided several comments regarding the terms used in the criteria. For example, terms like ''relevant,'' ''adequate,'' and ''appropriate'' were considered ambiguous. These terms are often used in critical appraisal tools of qualitative research [19]. Compared to reporting quality criteria, methodological quality criteria are more difficult to interpret because the reviewers need to judge whether the results that are reported can be trustworthy [43]. Also, criteria may be interpreted differently depending on the topic and context of the study.

The MMAT differs from other critical appraisal tools in several ways. To assess the quality of mixed methods studies, O'Cathain [11] suggested three different approaches: (a) generic research approach, (b) individual component approach, and (c) mixed methods approach. According to our review, the MMAT is the only tool that includes specific criteria for mixed methods studies [44]. With its five different sets of criteria, the MMAT uses a combination of individual component and mixed methods approaches. Other tools used in systematic mixed studies reviews approach critical appraisal differently. For example, Crowe and Sheppard [36] use a generic approach by proposing one set of criteria that could be applied to any design. Others, such as those from the Critical Appraisal

Skills Programme, Joanna Briggs Institute, and National Institute for Health and Clinical Excellence, propose one tool for each different study design (individual component approach). Also, some tools such as the Quality Assessment Tool for Studies with Diverse Designs (QATSDD) [45] use a combination of generic and individual component approaches, with generic criteria applicable to several designs and specific criteria for qualitative and quantitative studies.

In addition, the MMAT is distinct from the other tools in that it focuses on methodological quality criteria and consists of a small number of items. Similar to other risk of bias tools [46], the MMAT focuses on the core criteria that may hinder the validity of the findings of a study. Some criteria (such as information on ethical considerations), though essential in a research process, may have less impact on the validity of a study compared to other methodological criteria (such as appropriate measurement).

4.1. Strengths and limitations

Given that we found 15 reviews analyzing more than 500 critical appraisal tools, we considered that an overview of these reviews was an efficient approach to meet our objectives. Yet, it is likely that not all critical appraisal tools were included in the literature review because the search strategy did not include tools published in books and developed after 2015. For example, two recent literature reviews on tools for qualitative studies analyzed more than 100 tools [19,47]. Also, we limited our review to tools that had been validated or tested for reliability. Although it is possible that we did not identify all eligible critical appraisal tools, the pool of items we identified included over 75% criteria that were generic, reporting quality and duplicate. This suggests that our sample included the main criteria.

The number of experts on the three panels in Round-two ranged from 15 to 21. There is no rule regarding the required sample size for a Delphi. Some authors suggest a panel of 8 to 12 participants, whereas others recommended 300 to 500 [20]. One important factor to take into consideration when determining the size is the composition of the sample (homogeneous or heterogeneous). Usually, a smaller sample, such as 10 to 15 participants, is considered sufficient for homogeneous samples [20]. Similarly, there is no clear recommendation regarding the number of experts needed for content validation. Lynn [48] suggested that five experts could be sufficient. Polit, Beck, and Owen [49] recommended having 8 to 12 experts for the first round. Given this, because our samples were relatively homogenous in terms of experts' methodological expertise, their sizes may be considered acceptable.

Not all those who conduct systematic reviews are researchers with methodological expertise. Our study could have benefited from including such individuals in our panels of experts. For instance, the experience of health technology assessment practitioners or clinicians with experience in systematic reviews could have contributed to identifying relevant criteria to appraise. Future research and pilot testing of the MMAT could include this population.

The decision to use an agreement index threshold of 0.80 used in this study was arbitrary. There is no standard threshold for determining consensus in a Delphi study. Studies have used values varying from 0.50 to 0.80 [20]. In a previous study, it was found that criteria with an index of 0.78 or higher were indicative of good content validity [48]. Because the aim of this study was to identify core sets of criteria for validity content purpose, it was decided to use a high threshold.

Likert scales may have some limitations related to central tendency and desirability biases [50]. To limit this bias, we calculated frequencies (instead of means) and considered two ratings (very relevant and extremely relevant) to compute the agreement index.

5. Conclusion

The MMAT can facilitate the critical appraisal process in systematic mixed studies reviews by providing, within a single tool, methodological quality criteria for different designs. This modified e-Delphi sought experts' consensus on the methodological quality criteria of qualitative, survey, and mixed methods studies. The results led to replacing and clarifying the criteria of three of the five categories of studies in the MMAT and improving its content validity. Additional validation research on the MMAT is still needed, in particular, its discriminatory validity and inter-rater reliability.

CRediT authorship contribution statement

Quan Nha Hong: Conceptualization, Methodology, Investigation, Writing - original draft, Visualization. **Pierre Pluye:** Conceptualization, Methodology, Supervision, Resources, Writing - review & editing. **Sergi Fàbregues:** Validation, Writing - review & editing. **Gillian Bartlett:** Validation, Writing - review & editing. **Felicity Boardman:** Validation, Writing - review & editing. **Margaret Cargo:** Validation, Writing - review & editing. **Pierre Dagenais:** Validation, Writing - review & editing. **Marie-Pierre Gagnon:** Validation, Writing - review & editing. **Frances Griffiths:** Validation, Writing - review & editing. **Belinda Nicolau:** Validation, Writing - review & editing. **Alicia O'Cathain:** Validation, Writing - review & editing. **Marie-Claude Rousseau:** Validation, Writing - review & editing. **Isabelle Vedel:** Validation, Writing - review & editing.

Acknowledgments

The research team would like to acknowledge and sincerely thank all the e-Delphi panel experts for their contributions. Here are the names of the participants who

wished to be acknowledged: Lesley Andres (University of British Columbia, Canada); Theodore Bartholomew (Purdue University, United States); Pat Bazeley (Research Support/University of New South Wales, Australia); Jelke Bethlehem (Leiden University, Netherlands); Paul Biemer (RTI International, United States); Jaak Billiet (University of Leuven, Belgium); Felicity Bishop (University of Southampton, England); Jörg Blasius (University of Bonn, Germany); Hennie Boeije (University of Utrecht, Netherlands); Jonathan Burton (Understanding Society, England); Kathy Charmaz (Sonoma State University, United States); Benjamin Crabtree (The State University of New Jersey, United States); Elizabeth Creamer (Virginia Tech University, United States); Edith de Leeuw (University of Utrecht, Netherlands); Claire Durand (Université de Montréal, Canada); Joan Eakin (University of Toronto, Canada); Michèle Ernst Stähli (Université de Lausanne, Switzerland); Michael Fetters (University of Michigan Medical School, United States); Nigel Fielding (University of Surrey, England); Rory Fitzgerald (University of London, England); Floyd Fowler (University of Massachusetts, United States); Dawn Freshwater (University of Western Australia, Australia); Jennifer Greene (University of Illinois at Urbana-Champaign, United States); Christina Gringeri (University of Utah, United States); Greg Guest (FHI 360, United States); Timothy Guetterman (University of Michigan Medical School, United States); Muhammad Hadi (University of Leeds, England); Elizabeth Halcomb (University of Wollongong, United States); Carolyn Heinrich (Vanderbilt University, United States); Sharlene Hesse-Biber (Boston College, United States); Mieke Heyvaert (University of Leuven, Belgium); John Hitchcock (Indiana University Bloomington, United States); Nataliya Ivankova (University of Alabama at Birmingham, United States); Laura Johnson (Northern Illinois University, United States); Paul Lavrakas (University of Chicago, United States); Marilyn Lichtman (Virginia Tech University, United States); Geert Loosveldt (University of Leuven, Belgium); Peter Lynn (University of Essex, England); Mary Ellen Macdonald (McGill University, Canada); Claire Howell Major (University of Alabama, United States); Maria Mayan (University of Alberta, Canada); Sharan Merriam (University of Georgia, United States); José Molina-Azorín (University of Alicante, Spain); David Morgan (Portland State University, United States); Peter Nardi (Pitzer College, United States); Katrin Niglas (Tallinn University, Estonia); Karin Olson (University of Alberta, Canada); Antigoni Papadimitriou (Johns Hopkins University, United States); Michael Quinn Patton (Independent organizational development and program evaluation consultant, United States); Rogério Meireles Pinto (Columbia University School of Social Work, United States); Vicki Plano Clark (University of Cincinnati, United States); David Plowright (University of Hull, England); Blake Poland (University of Toronto, Canada); Rodney Reynolds (California Lutheran University, United States); Gretchen B. Rossman (University of Massachusetts Amherst, United States); Erin Ruel (Georgia State University, United States); Michael Saini (University of Toronto, Canada); Johnny Saldaña (Arizona State University, United States); Joanna Sale (Li Ka Shing Knowledge Institute, Canada); Karen Schifferdecker (Dartmouth College, United States); David Silverman (University of London, England); Ineke Stoop (Netherlands Institute for Social Research, Netherlands); Sally Thorne (University of British Columbia, Canada); Sarah Tracy (Arizona State University, United States); Frederick Wertz (Fordham University, United States). The authors gratefully acknowledge the sponsorship from the Method Development platform of the Québec SPOR SUPPORT Unit (#BRDV-CIHR-201-2014-05), the CIHR Doctoral Fellowship Award (#301011), and the FRSQ Senior Investigator Award (#29308).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.03.008.

References

[1] Bunn F, Trivedi D, Alderson P, Hamilton L, Martin A, Pinkney E, et al. The impact of Cochrane reviews: a mixed-methods evaluation of outputs from Cochrane review groups supported by the National Institute for health research. Health Technol Assess 2015;19:1−100.

[2] Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 1976;5(10):3−8.

[3] Cochrane AL. Effectiveness and efficiency: Random reflections on health services. London: Nuffield Provincial Hospitals Trust; 1972.

[4] Harden A, Gough D. Quality and relevance appraisal. In: Gough D, Oliver S, Thomas J, editors. An introduction to systematic reviews. London: SAGE Publications; 2012:153−78.

[5] Burls A. What is critical appraisal?. 2nd edn. Newmarket, UK: Hayward Medical Communications; 2009.

[6] Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: The Cochrane Collaboration and John Wiley & Sons Ltd; 2008:187−242.

[7] Wells K, Littell JH. Study quality assessment in systematic reviews of research on intervention effects. Res Soc Work Pract 2009;19(1):52−62.

[8] Wortman PM. Judging research quality. In: Cooper H, Hedges LV, editors. The handbook of research synthesis. New York: Russel Sage Foundation; 1994:97−109.

[9] Petticrew M, Roberts H. How to appraise the studies: an introduction to assessing study quality. In: Petticrew M, Roberts H, editors. Systematic reviews in the social sciences: A practical guide. Padstow, UK: Wiley-Blackwell; 2006:125−63.

[10] Pluye P, Robert E, Cargo M, Bartlett G, O'Cathain A, Griffiths F, et al. Proposal: A Mixed Methods Appraisal Tool for systematic mixed studies reviews 2011. Available at http://mixedmethodsappraisaltoolpublic.pbworks.com. Accessed November 15, 2013.

[11] O'Cathain A. Assessing the quality of mixed methods research: towards a comprehensive framework. In: Tashakkori A, Teddlie C, editors. Handbook of mixed methods in social and behavioral research. Thousand Oaks, CA: SAGE Publications; 2010:531−55.

[12] Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012;49(1):47−53.

[13] Souto RQ, Khanassov V, Hong QN, Bush PL, Vedel I, Pluye P. Systematic mixed studies reviews: updating results on the reliability and efficiency of the Mixed Methods Appraisal Tool. Int J Nurs Stud 2015;52(1):500−1.

[14] Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). J Eval Clin Pract 2018;24:459−67.

[15] Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. Psychol Assess 1995;7(3):238−47.

[16] Hong QN, Pluye P. A conceptual framework for critical appraisal in systematic mixed studies reviews. J Mix Methods Res 2018. https://doi.org/10.1177/1558689818770058. Advance online publication.

[17] Heyvaert M, Hannes K, Maes B, Onghena P. Critical appraisal of mixed methods studies. J Mix Methods Res 2013;7(4):302−27.

[18] Walsh D, Downe S. Appraising the quality of qualitative research. Midwifery 2006;22(2):108−19.

[19] Santiago-Delefosse M, Gavin A, Bruchez C, Roux P, Stephen S. Quality of qualitative research in the health sciences: analysis of the common criteria present in 58 assessment guidelines by expert users. Soc Sci Med 2016;148:142−51.

[20] Keeney S, Hasson F, McKenna H. The Delphi technique in nursing and health research. Chichester, UK: Wiley Online Library; 2011.

[21] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008−15.

[22] Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017;6(61):1−14.

[23] Baker J, Lovell K, Harris N. How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. Nurse Res 2006;14(1):59−70.

[24] von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. Technol Forecast Soc Change 2012;79(8):1525−36.

[25] Fàbregues S, Molina-Azorín JF. Addressing quality in mixed methods research: a review and recommendations for a future agenda. Qual Quant 2017;51(6):2847−63.

[26] Carroll C, Booth A. Quality assessment of qualitative evidence for systematic review and synthesis: is it meaningful, and if so, how should it be performed? Res Synth Methods 2015;6(2):149−54.

[27] Federal Committee on Statistical Methodology. Measuring and reporting sources of error in surveys. Washington DC: Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget; 2001.

[28] Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opin Q 2008;72(2):167−89.

[29] Fàbregues S, Paré M-H, Meneses J. Operationalizing and conceptualizing quality in mixed methods research: a multiple case study of the disciplines of education, nursing, psychology, and sociology. J Mix Methods Res 2018;. https://doi.org/10.1177/1558689817751774. Advance online publication.

[30] Pluye P, Garcia Bengoechea E, Granikov V, Kaur N, Tang DL. A world of possibilities in mixed methods: review of the combinations of strategies used to integrate the phases, results, and qualitative and quantitative data. Int J Mult Res Approaches 2018;10(1):41−56.

[31] Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs - Principles and practices. Health Serv Res 2013;48:2134−56.

[32] Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. Syst Rev 2017;6(204):1−9.

[33] Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Comparative Effectiveness Reviews; 2012.

[34] Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006;59:1249−56.

[35] Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. Chichester, UK: Wiley Online Library; 2008.

[36] Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011;64:79−89.

[37] Colle F, Rannou F, Revel M, Fermanian J, Poiraudeau S. Impact of quality scales on levels of evidence inferred from a systematic review of exercise therapy and low back pain. Arch Phys Med Rehabil 2002;83:1745−52.

[38] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94.

[39] Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 2015;12(10):e1001895.

[40] Davern M. Nonresponse rates are a problematic indicator of nonresponse bias in survey research. Health Serv Res 2013;48:905−12.

[41] Dillman DA, Phelps G, Tortora R, Swift K, Kohrell J, Berck J, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Soc Sci Res 2009;38(1):1−18.

[42] Bryman A. Integrating quantitative and qualitative research: how is it done? Qual Res 2006;6(1):97−113.

[43] Carroll C, Booth A, Lloyd-Jones M. Should we exclude inadequately reported studies from qualitative systematic reviews? An evaluation of sensitivity analyses in two case study reviews. Qual Health Res 2012;C22:1425−34.

[44] Pluye P. Critical appraisal tools for assessing the methodological quality of qualitative, quantitative and mixed methods studies included in systematic mixed studies reviews. J Eval Clin Pract 2013;19:722.

[45] Sirriyeh R, Lawton R, Gardner P, Armitage G. Reviewing studies with diverse designs: the development and evaluation of a new tool. J Eval Clin Pract 2012;18:746−52.

[46] Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011;343(d5928):1−9.

[47] Majid U, Vanstone M. Appraising qualitative research for evidence syntheses: a compendium of quality appraisal tools. Qual Health Res 2018;28:2115−31.

[48] Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382−6.

[49] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007;30:459−67.

[50] Jamieson S. Likert scales: how to (ab)use them. Med Educ 2004;38(12):1217−8.

Appendix

Appendix 1. Mixed Methods Appraisal Tool (MMAT) version 2018

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?				
	S2. Do the collected data allow to address the research questions?				
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
1. Qualitative	1.1. Is the qualitative approach appropriate to answer the research question?				
	1.2. Are the qualitative data collection methods adequate to address the research question?				
	1.3. Are the findings adequately derived from the data?				
	1.4. Is the interpretation of results sufficiently substantiated by data?				
	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?				
2. Quantitative randomized controlled trials	2.1. Is randomization appropriately performed?				
	2.2. Are the groups comparable at baseline?				
	2.3. Are there complete outcome data?				
	2.4. Are outcome assessors blinded to the intervention provided?				
	2.5 Did the participants adhere to the assigned intervention?				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?				
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?				
	3.3. Are there complete outcome data?				
	3.4. Are the confounders accounted for in the design and analysis?				
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?				
4. Quantitative descriptive	4.1. Is the sampling strategy relevant to address the research question?				
	4.2. Is the sample representative of the target population?				
	4.3. Are the measurements appropriate?				
	4.4. Is the risk of nonresponse bias low?				
	4.5. Is the statistical analysis appropriate to answer the research question?				
5. Mixed methods	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?				
	5.2. Are the different components of the study effectively integrated to answer the research question?				
	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?				
	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?				
	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?				