

Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT)

Quan Nha Hong, OT, MSc, PhD candidate, Department of Family Medicine, McGill University, 5858 Côte-des-Neiges, Suite 300, Montréal, QC, H3S 1Z1, Canada, quan.nha.hong@mail.mcgill.ca

Araceli Gonzalez-Reyes, MSc, PhD candidate, Department of Family Medicine, McGill University, 5858 Côte-des-Neiges, Suite 300, Montréal, QC, H3S 1Z1, Canada, araceli.gonzalezreyes@mail.mcgill.ca

Pierre Pluye, MD, PhD, Full professor, Department of Family Medicine, McGill University, 5858 Côte-des-Neiges, Suite 300, Montréal, QC, H3S 1Z1, Canada, pierre.pluye@mcgill.ca

Corresponding Author:

Pierre Pluye, Department of Family Medicine, McGill University, 5858 Côte-des-Neiges, Suite 300, Montréal, QC, Canada, H3S 1Z1; Tel: 1-514-398-8483, Fax: 1-514-398-4202.

Email address: pierre.pluye@mcgill.ca

Running title: Usefulness of the Mixed Methods Appraisal Tool

Keywords: qualitative research, quality appraisal, mixed studies reviews, systematic review, usefulness, usability, utility.

ABSTRACT

Rationale, aims and objectives: Systematic reviews combining qualitative, quantitative, and/or mixed methods studies are increasingly popular due to their potential for addressing complex interventions and phenomena, specifically for assessing and improving clinical practice. A major challenge encountered with this type of review is the appraisal of the quality of individual studies given the heterogeneity of the study designs. The Mixed Methods Appraisal Tool (MMAT) was developed to help overcome this challenge. The aim of this study was to explore the usefulness of the MMAT by seeking the views and experiences of researchers who have used it.

Methods: We conducted a qualitative descriptive study using semi-structured interviews with MMAT users. A purposeful sample was drawn from the researchers who had previously contacted the developer of the MMAT, and those who have published a systematic review for which they had used the MMAT. All interviews were transcribed verbatim and analyzed by two coders using thematic analysis.

Results: Twenty participants from eight countries were interviewed. Thirteen themes were identified and grouped into the two dimensions of usefulness, i.e., utility and usability. The themes related to utility concerned the coverage, completeness, flexibility, and other utilities of the tool. Those regarding usability were related to the learnability, efficiency, satisfaction and errors that could be made due to difficulties understanding or selecting the items to appraise.

Conclusions: On the basis of the results of this study, we make several recommendations for improving the MMAT. This will contribute to greater usefulness of the MMAT.

INTRODUCTION

Systematic reviews combining qualitative, quantitative, and/or mixed methods studies, are more and more popular due to their potential for addressing complex evaluation questions that matter in clinical practice^{1,2}. Indeed, including different types of studies in a review can provide a richer understanding of the impact of contextual factors, help focusing on outcomes that are important for patients, and explore the diversity of effect across studies³. These reviews have various labels such as systematic mixed studies reviews⁴, mixed methods research synthesis⁵, and integrative review⁶. The first label refers to combining qualitative, quantitative and/or mixed methods studies while the second one can also refer to combining qualitative and quantitative methods (such as thematic synthesis and meta-analysis)⁷. Hereinafter, we will use the term ‘systematic mixed studies reviews’ to designate this type of review. While they are increasing popular⁷, these reviews present several challenges given the heterogeneous nature of study designs, including the critical appraisal of the quality of individual studies. Critical appraisal is a core step of systematic reviews and consists of a systematic and careful examination of studies to ensure they are trustworthy^{8,9}.

Critical appraisal tools have been developed to formalize the quality appraisal process and ensure it is done in a systematic, transparent, and reproducible manner¹⁰. A large variety of these tools exists and are, for most part, checklists and scales of quality appraisal items¹¹. For example, authors of literature reviews have identified 94 tools for randomized controlled trials (RCT)¹², 194 for nonrandomized studies¹³, 13 for mixed methods studies¹⁴, and 58 for qualitative research¹⁵. The wide variety makes it difficult for reviewers to choose the most appropriate one(s). This is particularly true for systematic mixed studies reviews since the heterogeneity in the designs of the included studies requires that reviewers search for, select, and learn how to use several tools. Also, there is a lack of agreement regarding the most appropriate critical appraisal tools and approaches to use¹¹⁻¹³. Many tools were not developed using rigorous development process including sound validation and reliability testing¹⁶⁻¹⁸. To address this, Whiting et al.¹⁹ recently proposed a framework for developing quality assessment tools, which includes three key stages: initial steps (including identifying needs and scope for a new tool), tool development, and dissemination.

The Mixed Methods Appraisal Tool (MMAT) allows for the critical appraisal of quantitative, qualitative, and mixed methods studies and was developed to address the challenges

of critical appraisal in systematic mixed studies review. The MMAT is rooted in a literature review on systematic mixed studies reviews conducted in 2006⁴. To provide proof of concept of the feasibility of the MMAT, the research team conducted a pilot study and subsequent studies of interrater reliability. These studies showed that it is relevant to researchers and decision/policymakers and feasible for them to use²⁰, and that there is a variability of agreement of the items ranging from poor to perfect and a need for further testing and refinement of this tool^{20, 21}. To further the development and testing of the MMAT, more research is needed with researchers who had used this tool.

Since its development, the MMAT has been cited in more than one hundred systematic reviews, and its website²² has been visited more than 20,000 times. This widespread use made it possible to explore the views and experiences of researchers who have used the MMAT and were not directly involved in its initial development (hereinafter “MMAT users”). Our research question was: What are the views and experiences of researchers regarding the use of the MMAT? The results of this study with users contributed to identifying the key areas for improvement that is required in the MMAT.

METHODS

A qualitative descriptive method^{23, 24} was employed with MMAT users. This method fits well with the aim of this project that focused on describing the experience of MMAT users. This method stays close to the data and focuses on reporting the manifest content of data, rather than being highly interpretive and conceptual^{23, 24}. Qualitative description is appropriate in mixed methods research for the development and refinement of questionnaires or interventions²⁵.

Description of the Mixed Methods Appraisal Tool (MMAT)

The latest version of the MMAT (version 2011) includes two screening questions and 19 items for appraising the methodological quality of five categories of studies: qualitative studies (4 items), randomized controlled trials (4 items), nonrandomized studies (4 items), quantitative descriptive studies (4 items), and mixed methods studies (3 items). The screening questions are used to exclude non-empirical studies from the appraisal stage, i.e., research that is not based on experience (e.g., observation, experiment, or simulation) such as reviews and theoretical papers²⁶. The MMAT was conceived so that one set of items can be used when appraising a

qualitative or quantitative study. When appraising mixed methods studies, three sets of items are assessed: the qualitative set, a quantitative set (either, the randomized controlled trial, nonrandomized studies, or the quantitative descriptive studies), and the mixed methods set. Each item is rated on a categorical scale (yes, no, and cannot tell) and the number of items rated “yes” are counted to provide an overall score (see supplementary file). The MMAT is available online (<http://mixedmethodsappraisaltoolpublic.pbworks.com>) and comes with a user manual (tutorial) in which each item is described, and examples and references are provided. For each category of studies, examples of common study designs are provided (see supplementary file).

Study Participants

A purposeful sample of researchers with experience using the MMAT was generated by two means. First, forward citation tracking of three papers on the MMAT^{4, 20, 22} was performed on September 6, 2015 in Google Scholar. These references had been cited, respectively, 51, 156, and 54 times. From these citations, we selected the systematic reviews published after 2011 (year of the latest version of the MMAT) that included more than 10 studies and collected the name and email address of the first authors. Second, the primary developer of the MMAT had a list of 81 researchers who had contacted him over the years requesting permission to use it for research or training purposes, clarification on how to use it, or requesting for the latest version. A maximum variation sampling was used to account for the different institutions, countries, and occupations of these researchers. An email was sent to 72 researchers inviting them to participate in an interview in English or French regarding their experience using the MMAT.

Data Collection

We conducted semi-structured interviews with MMAT users either through Skype or Go-To-Meeting. During the interview, a semi-structured guide was used to collect information pertaining to their (a) research experience (e.g., fields of interest, number of years of research experience, research methods experience, and occupation), and (b) experience using the MMAT (e.g., number of papers appraised using the MMAT, study designs of the papers appraised, perceived utility of the MMAT). The interview guide was developed to elicit MMAT users’ perspectives and experiences with different parts of the tool, i.e., the items, the scale, the tutorial, and the five study design sets. Five questions were posed: (a) What do you like about the MMAT

and why?, (b) What do you dislike about the MMAT and why?, (c) Did you encounter any problems when using the MMAT?, (d) Did you make any changes to the tool during your project?, and (e) Were you able to use the MMAT to appraise all the papers included in your reviews? The interviews were recorded using a digital recorder.

The interview guide was piloted with three students who had used the MMAT in their master's research. This pilot test aimed to verify the clarity of the questions and their order, to estimate the time of the interview, and to test different communication media (phone, Skype, and in-person) and the recording quality.

Data Analysis

A professional transcriber transcribed the interviews and the interviewer checked the verbatim transcripts for accuracy. The transcripts were analyzed using thematic analysis²⁷. Two coders independently coded the transcripts using a specialized software program (NVivo 11). Initially, they used open coding, reading and re-reading the transcripts to generate a preliminary list of codes. After analyzing three interviews, the two coders met to compare and discuss their codes and establish a codebook. This process was iterative and repeated until no substantive new codes were identified. The codes were then analyzed and combined into meaningful groups to identify initial themes. At this stage, the themes were grouped into three broad categories: strengths of the MMAT, difficulties encountered when using the MMAT, and changes made or suggested in the MMAT. Once the themes were identified, the team met to discuss how to organize them coherently and meaningfully. Discussions among the team led to using the framework on system acceptability to organize the themes (Figure 1). Developed in the field of human-computer interaction, this framework presents the main dimensions required to ensure that a system is good enough to satisfy the users' needs and requirements²⁸. Within this framework, a system overall acceptability is composed of its social and practical acceptability. To analyze the practical acceptability of a system, several dimensions can be considered such as its cost, reliability and usefulness. In this study, we focused on the usefulness dimension that is defined as whether the system can achieve its desired goal²⁸. We considered that the MMAT is a system that users will use to achieve the intended goal of appraising the quality of qualitative, quantitative, and mixed methods studies. All the themes identified in the open coding were interpreted using, and grouped into, the dimensions of usefulness in this framework, i.e., utility

and usability (Figure 1).

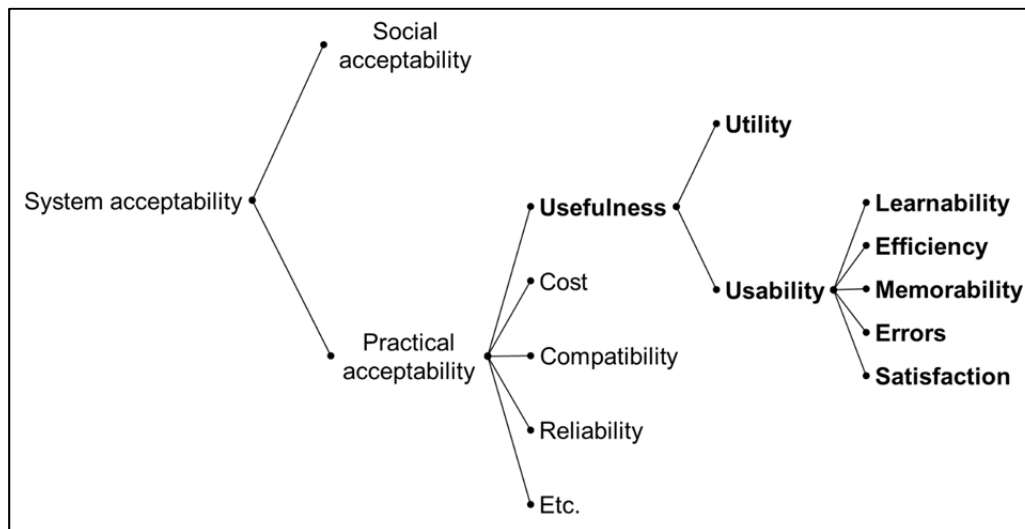


Figure 1. Framework on system acceptability, Reprinted from Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994, with permission from Elsevier.

Ethical Considerations

This study was approved by the Institutional Review Board (IRB) of the Faculty of Medicine of McGill University (project # A05-E26-15B). All participants completed a consent form prior to the interview. Participants were numbered and no identifying information was presented in the data file used for the analysis.

RESULTS

A total of 72 invitation emails were sent between November 2015 and March 2016, of which 20 resulted in interviews. The reasons for non-participation were: did not respond to the invitation email (n=42), had invalid email address or out-of-office message (n=4), had not used the MMAT (n=3), was not available during the period of the interviews (n=1), was not interested (n=1), and used the MMAT too long ago to remember (n=1). The interviews were conducted in English (n=16) or French (n=4), and lasted between 21 and 48 minutes.

The 20 participants were affiliated with institutions from eight different countries. They were mostly female (n=17) and affiliated with a university (n=19). Their research areas were predominantly in health sciences (including nursing, public health, global health, community health, palliative care, primary care, cardiovascular, oncology, and gerontology). Nearly half of the participants were doctoral candidates. Most were mixed methods researchers (n= 9); whereas

the others identified themselves as primarily qualitative (n=5) or quantitative (n=6) researchers (Table 1). With the exception of one participant who used the MMAT in a journal club, all had used it in a systematic review. Participants used the MMAT results to describe the quality of included studies (n=14), exclude studies from the review (n=3), justify the quality criteria extracted from studies (n=1), make recommendations (n=1), and compare with the appraisal of other critical appraisal tools (n=1).

Table 1. Profile of participants

Characteristics		Number
Countries	Australia	4
	Canada	5
	Denmark	1
	England	3
	France	1
	Netherlands	2
	New Zealand	1
	United States	3
Occupation	Doctoral candidate	9
	Post-doctoral fellow	4
	Professor/lecturer	5
	Research associate	1
	Librarian	1
Gender	Female	17
	Male	3
Setting	Public health agency	1
	University	19
Main research methodology used	Qualitative	5
	Quantitative	6
	Mixed methods	9
Main research areas of interest	Architecture	1
	Education	1
	Health sciences	14
	Information sciences	1
	Physical activity	2
	Psychology	1
Year of experience in research, mean (SD)		6.7 (3.7) years

A total of 13 themes were identified and grouped into the two dimensions of usefulness, i.e., utility and usability (Figure 1).

Utility

Utility is defined as whether or not the tool can function as needed²⁸. Five themes were found regarding the utility of the MMAT; two addressed its coverage, and one each for completeness, flexibility, and other utilities.

Coverage

Two themes were related with the scope of designs covered by the MMAT.

Theme 1 – Comprehensive tool: The MMAT users appreciated that the tool can be applied to several study designs (qualitative, quantitative, and mixed methods studies):

The thing that I liked about it it's an all in one package. [...] There's so much to write, there's so much to analyze, if you have 2 different or 3 different tools to use. I can see that the development of this tool was also based on previous work of critical appraisal and all of those. But for me, what is very... what I really liked the most about it, it's there, it's all in one. You can use it... yeah... You don't have to use any other tool. (P16)

Theme 2 - Study designs that could not be appraised with the tool: Some MMAT users mentioned that the items in the MMAT were less relevant for some study designs such as cost-effectiveness studies, political analysis, transcultural adaptation, and pragmatic trials:

Like studies in political science, political analysis, policy development process, they did not really fit with the MMAT. A second type of studies that I had difficulty assessing with the MMAT were studies in economics, cost-effectiveness studies. (P04)

Completeness

One theme addressed concerns about the completeness of the tool. The completeness refers to the degree to which all important items to appraise the quality of studies are included in the MMAT.

Theme 3 - Concerns about completeness of the tool: Because the MMAT includes four items for each research design set, MMAT users were concerned that the tool might be 'too simple', 'superficial', 'global', and would not discriminate 'good' and 'bad' studies. Some MMAT users mentioned that items were missing in the tool such as those concerning conflict of interest,

quality of reporting, confounding variables, selective reporting bias, sample size, external validity, theoretical underpinnings, publication bias, triangulation, data analysis, and ethics.

So my concern, initially when I first started to use it and when comparing with other types of appraisal tools, I was afraid that it might be missing some appraisal items. At the time it had been... the pilot study had been done for validation, so that was reassuring, but at the same time that would have been just concerns about completeness. (P18)

Flexibility

One theme pertained to the need to adapt the MMAT. We interpreted this to be about the flexibility of the tool, which refers to its ability to be modified based on the research topic or study design.

Theme 4 – Need to adapt the tool to the topic of the review: Some users suggested having a more flexible tool that could be tailored to the topic of their review. For example, they suggested providing more weight to certain items or adding optional items they judged important in their field. Also, some MMAT users questioned the utility of the two screening questions and suggested that they be removed when the selection criteria are limited to empirical studies.

Moreover, they suggested having cut-off values in the items that could be adapted to their field.

And also in the observational ones, we wanted to be able to discriminate or give a bit of a better weighting to perspective of longitudinal studies. So within the justification of measurements, we also rated it high, we also gave an extra point if it was longitudinal perspective compared to cross-sectional. (P17)

Other utilities

In addition to appraising the quality of studies, some users mentioned that the MMAT can have additional utility.

Theme 5 – Educational tool: The MMAT users liked that the tool was helpful to learn about study designs and that it was a relevant resource for graduate students:

And it's a really nice resource for students particularly, because we want to encourage them to think broadly when they think of systematic review and not to just think of the quantitative systematic review of intervention studies or the meta-analysis kind of reviews. And that gets really overwhelming. So this tool kind of consolidates a lot of ways of thinking about the quality of your studies into a single document that's useful for them to think through. (P08)

Usability

Usability is defined as how well users can use the tool²⁸. Compared with utility where no attribute is specified in the system acceptability framework, five usability attributes are defined: learnability, efficiency, memorability, errors and satisfaction (Figure 1)²⁸. In this study, eight themes on usability were found and were related to four of these attributes.

Learnability

Learnability refers to how the tool is easy to learn²⁸. Two themes were found on this attribute.

Theme 6 – Easy to use: The MMAT users liked that the tool was easy to understand, rate, and use:

[...] it was really clearly explained how you can include and exclude, how you're supposed to evaluate the studies, it was really well laid out. Easy for someone who's never done this kind of thing before to follow. The instructions are really good. (P19)

Theme 7 – Improvements needed in the tutorial: Several comments were made on the tutorial. The MMAT users found the tutorial helpful to refer to. They appreciated the list of study designs and the explanations of the items. However, they mentioned that some explanations provided did not match the items. Some MMAT users suggested expanding the study designs list to include, for instance, interpretive description, comparative studies, and survey. They also suggested adding information in the tutorial to facilitate the use of the tool, such as a title page, the explicit purpose of the MMAT, and an algorithm. Many mentioned that it was unclear how to score “cannot tell” response category and some suggested modifying the scale. Moreover, MMAT users suggested adding more examples of how to rate items, and clarifying how to compute an overall score and how to present the results of the appraisal:

The left-hand box is really useful. That's good because it helps you to classify the type of qualitative, what you've got. The right-hand side, where you're asking the questions, possibly give more specific examples maybe of what there is there. Because in all cases, I would say that would be relevant really. Like on the left-hand side definitely that's fine, I would leave it there. But maybe add some more examples on the right-hand side. (P21)

Efficiency

Efficiency is defined as allowing for a high level of performance once the users have learned to use the tool²⁸. One theme addressed this attribute.

Theme 8 – Short and quick: The MMAT users liked that the tool was simple, short, and allowed

for completing study appraisal quickly:

I liked that it's simple and it's not too long. It's not an enormous task to go through. It's very clear to see which bits are going to be relevant to what I need. I can just go straight in there and see which areas I need to look at. (P14)

Errors

Errors are defined as actions that do not accomplish the intended goal²⁸. Given the goal of the MMAT is to appraise the quality of qualitative, quantitative, and mixed methods studies, we included in this attribute two themes on difficulties understanding the items or selecting of the items to appraise.

Theme 9 - Items not clear or difficult to judge: The MMAT users provided comments on items that were difficult to understand and rate. Four subthemes were identified.

Subtheme 9.1- Qualitative and mixed methods studies subject to interpretation: Several comments were made concerning items in the qualitative and mixed methods studies item sets that were considered more difficult to judge, more open to interpretation or less precise compared with the quantitative study designs items.

There was some... and I think this is acknowledged in the template, some of the criteria were a little bit difficult to interpret, particularly around kind of the qualitative items about researchers' influence and the context, which were difficult to establish.[...] The mixed methods was similar to the qualitative ones in that they were a bit open to interpretation compared with the quantitative items. (P15)

Subtheme 9.2- Several concepts in one item: MMAT users commented on the fact that some items include several concepts and suggested clarifying or modifying these items.

I think the ideas are quite clear. However, there are several concepts in the same question. So here, I think that was what I found difficult. Take question 3.2: 'are measurements appropriate regarding the exposure, control', etcetera. You see that in the parentheses there are a lot of concepts and each of these concepts could be a sub-question. (P11)

Subtheme 9.3 – Missing information in papers: MMAT users pointed out that some items were considered more difficult to judge because of missing information in the papers appraised.

But 'is appropriate consideration given to how findings relate to the context?', that's very hard. And then 1.4 'is consideration given to how the researchers' influence or the

interaction with the participants?’. Because of this word limit of publications and because qualitative... When you write a qualitative paper, you’re already struggling for space, because most health science journals only allow you 3000 words. It’s already a struggle to put in your citations and everything and everything counts, so I don’t... I can’t remember I have read a paper that goes into detail about how the researcher might have influenced the findings etcetera. (P07)

Subtheme 9.4 – Unclear distinction between some items: MMAT users mentioned that the distinction between some items is subtle.

The one that I probably used least often and the one that I had the most questions about, - but again, I’m not using it all that often - is the RCT, the difference between the complete outcome data of 80% and the low withdrawal rate of 20%. That’s a very fine line in my mind of what’s the differentiation. (P13)

Theme 10 - Difficulty classifying the studies: MMAT users mentioned that they had difficulty deciding if they should use the nonrandomized or the descriptive sets:

One of the things... quantitative nonrandomized... quantitative descriptive... We had problems trying to classify some of the studies. We didn’t have specific enough, sufficient details for you to be able to tell what type of study it is. So we ended up classifying the majority of studies as quantitative descriptive mainly because we didn’t have sufficient information from the studies themselves (P09).

Satisfaction

Satisfaction refers to how pleasant the tool is to use²⁸. Three themes were related to this attribute.

Theme 11 - Accessible online: The MMAT users liked that the tool was available online:

Another thing that I really liked about the tool is that it’s online and everybody can get access to it. [...] So when people ask me about that, I said “I can send you a link but it’s right there online, you can just go in and look at it”. And it’s really really helpful for people. (P05)

Theme 12 - Website not user-friendly: The MMAT users provided comments on the navigation of the website:

I do remember being on your website and your website might be just a little bit tricky to navigate. (P06)

Theme 13 – Missing rating sheet: MMAT users proposed providing a rating sheet, such as an

Excel document, that could be used to compute the ratings and calculate an overall score:

The only thing was that it was not available in a document that you can write in.[...] Yes like the Excel sheet I showed you. I don't know if that would be helpful. I just made it myself. (P10)

DISCUSSION

The development of the MMAT followed the framework for developing quality assessment tools¹⁹: initial steps (e.g., identify a need for an appraisal tool for systematic mixed studies studies), tool development (e.g., literature review, pilot testing, reliability testing), and dissemination (e.g., workshops, website, publications). However, this process is not linear and should include feedback loops to revise and refine the tool. To contribute to the revision of the MMAT, we explored the views and experiences of researchers who have used it. We identified 13 themes and classified them according to the dimensions of usefulness (utility and usability) as suggested by Nielsen²⁸. Table 2 presents a summary of the themes. Regarding utility, our results pertain to the coverage, completeness, flexibility, and other utility of the MMAT. In term of usability, our findings point to issues of learnability, efficiency, errors, and user satisfaction. Some themes suggest potential areas for improvement in the MMAT (see * in Table 2).

The MMAT users appreciated that the tool was easy to use, comprehensive, quick, short, and accessible online. These themes are considered strengths of the MMAT that should be maintained in subsequent revision of the MMAT. Having pre-defined items can be helpful to ensure that the key methodological aspects are examined in a systematic and transparent manner using a common approach for all included studies¹⁰. Since systematic mixed studies reviews can include a wide range of study designs, these tools can be particularly appealing to graduate students and researchers who are unfamiliar with certain study designs.

The results of this study can contribute to improve the ecological validity of the MMAT. Ecological validity is a subset to external validity and refers to the transferability of findings from an experimental context to the real-world environment^{29, 30}. Interviewing other users that were not involved with the development of the MMAT can provide different and some more impartial views of the MMAT.

Table 2. Themes identified in this study

Dimensions	Attributes	Themes*
Utility	Coverage	1 – Comprehensive tool
		2 – Study designs that could not be appraised with the tool*
	Completeness	3 – Concerns about the completeness of the tool*
	Flexibility	4 – Need to adapt the tool to the topic of the review*
	Other utility	5 – Educational tool
Usability	Learnability	6 – Easy to use
		7 – Improvement needed in the tutorial*
	Efficiency	8 – Short and quick
		9 – Items not clear or difficult to judge*
		9.1 – Qualitative and mixed methods studies subject to interpretation
		9.2 – Several concepts in one item
	Errors	9.3 – Missing information in papers
		9.4 – Unclear distinction between some items
		10 – Difficulty classifying the studies*
	Satisfaction	11 – Accessible online
		12 – Website not user-friendly*
		13 – Missing rating sheet*

* Themes suggesting potential areas for improvement

Recommendations for Improving the MMAT

On the basis of our results, six recommendations can be put forward for the MMAT.

First, the MMAT includes criteria for five broad categories of study designs and specific criteria for each design. Yet, our results show that choosing items is difficult for some studies, in particular for cross-sectional and single group studies. This difficulty could be addressed by clarifying the study design categories and adding a selection algorithm such as those developed and tested in Hartling et al.³¹, and Seo et al.³² for classifying quantitative study designs.

Second, the MMAT is focused on appraising methodological quality. In this study, the MMAT users underscored an important usability issue: poor reporting hinders the appraisal of some MMAT items. Inadequate reporting precludes adequate appraisal of how a study was conducted and its results³³. Moreover, lack of reporting about a methodological criterion does not mean it was not met in the study^{34, 35}. To address this issue, the MMAT has a ‘Cannot tell’ response category and it is suggested to contact the researchers to obtain additional information. This approach has been critiqued since it can lead to risk of overly positive answers (i.e.,

tendencies of providing positive answers that do not necessarily reflect the reality of a study)³⁶. Given that the reliability of the information provided may be questionable, some have recommended limiting the appraisal to published material and matching the quality of reporting with the level of information needed to appraise the methodological quality of a study³⁷. This recommendation is an avenue to explore for the revised version of the MMAT. Items to include in the MMAT could be chosen on the basis of information that is typically reported. Another potential avenue is a two-step approach where inadequately reported papers are excluded on the basis of an initial reporting quality appraisal, and, methodological quality of the remaining papers is subsequently appraised³⁸. Carroll et al.³⁸ tested this approach and found that excluding inadequately reported papers does not influence the overall results of the synthesis in qualitative systematic reviews, although it might lead to exclusion of particular disciplines/perspectives. In a recent review, Verhage and Boels³⁹ concurred with Carroll et al.³⁸, but mentioned that, although the exclusion of inadequately reported papers does not affect the number and nature of the themes identified, it may influence the degree of nuance and the richness of the themes.

Third, the current version of the MMAT has four items per category of study design, which is few compared to other critical appraisal tools. Although our results show that the short and comprehensive nature of the MMAT is appreciated, they also indicate concerns about its utility due to its lack of completeness and missing items. The MMAT developers chose to focus on efficiency, including only the most important items for judging the methodological quality of a study. Yet, in tool development, it is necessary to ensure that the tool adequately covers the construct is meant to assess (i.e., the methodological quality of studies in the case of the MMAT). This is related to the content validity of the tool⁴⁰ and will need to be further explored with methodological experts.

Fourth, our results suggest that the qualitative and mixed methods studies items are difficult to judge. These items were considered more subject to interpretation and less precise than the quantitative items. Several reasons could explain this difficulty such as the lack of reporting (e.g., unclear description and lack of details) precluding a proper appraisal, and the unfamiliarity of the reviewers with these types of studies. Also, in the MMAT, only one set of items were developed for qualitative and mixed methods studies while there are three different sets of items for quantitative studies (RCT, nonrandomized, and descriptive). There is a need to provide more explanations and examples about how to interpret and rate these items in the

MMAT. Also, further studies could explore the need to add items regarding specific qualitative approaches (e.g., qualitative descriptive, grounded theory, ethnography, and phenomenology).

Fifth, our results suggest making the MMAT more flexible by, for instance, adding optional, weighting items or modifying the cut-off values when judged necessary by the reviewers. This could improve the utility of the tool and help tailor it to the needs of the users. This is in line with Santiago-Delefosse et al.¹⁵ who promote a flexible list of criteria for qualitative research based on their study with 46 participants. They found that consensus can be reached only for general criteria and that there was a lack of consensus on the definition of criteria and their weights. In addition to having core criteria for each design, the MMAT could include a list of validated items from which the researchers can choose to meet the specific needs of their review.

Sixth, the users' satisfaction when using a tool is another important usability issue that needs to be considered when developing a critical appraisal tool. Users who are pleased with the tool tend to recommend it to others²⁸. Complementary materials such as a user manual or website can enhance users' satisfaction. On the basis of our results, concrete improvements to enhance users' satisfaction with the MMAT should be made such as improving the website navigation, providing more examples of rating in the tutorial, and adding a rating sheet.

Strengths and Limitations

We interviewed 20 MMAT users. Similar themes were mentioned by the MMAT users and data saturation was reached; further interviews would probably not have added new information to the overall results⁴¹. After the 8th interview, no new code emerged. The addition of interviews helped to provide more information on the themes. While our sample was heterogeneous with participants from several countries, working on a wide range of research topics mainly in health care, and having different expertise, almost all participants worked in university settings. Other potential MMAT users, such as health technology assessment professionals, were not reached. Also, nearly half of the participants were doctoral candidates, which can be representative of the main MMAT users. Indeed, systematic review is a method increasingly used at the graduate level. Some even suggest that systematic reviews be mandatory in doctoral programs⁴².

Two authors of this study are familiar with the MMAT. The interviews were performed

by the first author who was a doctoral candidate at the time of this study. She has gained experience with the MMAT one year prior to the interviews by collaborating as a second reviewer on systematic reviews. The last author is one of the developers of the MMAT and has been working on this tool since 2006. Their preconceptions of the MMAT could have influenced the interviews and analyses. Care was taken to make sure the data collected and analyzed represent the experience of the MMAT users such as involving a second coder that was not familiar with the MMAT, having independent coding, and developing a codebook. The coders did not encounter difficulties in reaching a consensus since the level of interpretation of data was low (analysis of the manifest content of interviews).

CONCLUSION

As systematic mixed studies reviews are gaining in popularity, appraisal tools that can be used to assess different study designs are needed. This study with MMAT users is a first important step in the improvement of its usefulness. The 13 themes identified and grouped into the system acceptability framework may be useful for developers of other critical appraisal tools.

ACKNOWLEDGEMENTS

The research team is grateful to the 20 participants who generously shared their time, perspectives and experiences. Also, the authors would like to thank the two anonymous reviewers and Dr. Paula Bush, Academic Associate at McGill University, for their constructive comments that helped to improve and clarify this manuscript.

Quan Nha Hong holds a Doctoral Fellowship Award from the Canadian Institutes of Health Research (CIHR) (#301011). Araceli Gonzalez-Reyes holds a Doctoral Fellowship Award from the Fonds de recherche du Québec – Santé (FRQS) (#28715). Pierre Pluye holds a Senior Investigator Award from the FRQS (#29308).

REFERENCES

1. Heyvaert M, Maes B, Onghena P. Mixed methods research synthesis: Definition, framework, and potential. *Qual Quant*. 2013;47(2),659-676. doi:10.1007/s11135-011-9538-6.
2. Pluye P, Hong QN. Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annu Rev Public Health*. 2014;35,29-45. doi:10.1146/annurev-publhealth-032013-182440.
3. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Integrative approaches to qualitative and quantitative evidence. London, UK: Health Development Agency; 2004.
4. Pluye P, Gagnon MP, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *Int J Nurs Stud*. 2009;46(4),529-546. doi:10.1016/j.ijnurstu.2009.01.009.
5. Heyvaert M, Hannes K, Onghena P. Using Mixed Methods Research Synthesis for Literature Reviews: The Mixed Methods Research Synthesis Approach. Thousand Oaks, CA: Sage Publications; 2016.
6. Whitemore R, Knafl K. The integrative review: Updated methodology. *J Adv Nurs*. 2005;52(5),546-553. doi:10.1111/j.1365-2648.2005.03621.x.
7. 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(1),61. doi:10.1186/s13643-017-0454-2.
8. Harden A, Gough D. Quality and relevance appraisal. In Gough D, Oliver S, Thomas J, eds., *An Introduction to Systematic Reviews* London, UK: Sage Publications; 2012:153-178.
9. Burls A. *What is Critical Appraisal?* Newmarket, UK: Hayward Medical Communications; 2009.
10. Petticrew M, Roberts H. How to appraise the studies: an introduction to assessing study quality. In Petticrew M, Roberts H, eds., *Systematic Reviews in the Social Sciences: A Practical Guide* Padstow, UK: Wiley-Blackwell; 2006:125-163.
11. West SL, King V, Carey TS, et al. *Systems to Rate the Strength of Scientific Evidence*. Rockville, MD: Agency for Healthcare Research and Quality; 2002.
12. Bai A, Shukla VK, Bak G, Wells G. *Quality Assessment Tools Project Report*. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012.

13. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess.* 2003;7(27),iii-x. doi:10.3310/hta7270.
14. Heyvaert M, Hannes K, Maes B, Onghena P. Critical appraisal of mixed methods studies. *J Mix Methods Res.* 2013;7(4),302-327. doi:10.1177/1558689813479449.
15. 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;148,142-151. doi:10.1016/j.socscimed.2015.11.007.
16. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol.* 2004;4(22). doi:10.1186/1471-2288-4-22.
17. Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *J Clin Epidemiol.* 2011;64(1),79-89. doi:10.1016/j.jclinepi.2010.02.008.
18. Sanderson S, Tatt ID, Higgins JP, Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *Int J Epidemiol.* 2007;36(3),666-676. doi:10.1093/ije/dym018.
19. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. *Syst Rev.* 2017;6(1),204. doi:10.1186/s13643-017-0604-6.
20. Pace R, Pluye P, Bartlett G, 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. doi:10.1016/j.ijnurstu.2011.07.002.
21. 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),500-501. doi:10.1016/j.ijnurstu.2014.08.010.
22. Pluye P, Robert E, Cargo M, et al. Proposal: A Mixed Methods Appraisal Tool for systematic mixed studies reviews. In: Archived by WebCite® at <http://www.webcitation.org/5tTRTc9yJ> 2011.
23. Sandelowski M. Focus on research methods-whatever happened to qualitative description? *Res Nurs Health.* 2000;23(4),334-340. doi:10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G.

24. Sandelowski M. What's in a name? Qualitative description revisited. *Res Nurs Health*. 2010;33(1),77-84. doi:10.1002/nur.20362.
25. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description—the poor cousin of health research? *BMC Med Res Methodol*. 2009;9(1),52.
26. Porta MS, Greenland S, Hernán M, dos Santos Silva I, Last JM. *A Dictionary of Epidemiology*. New York: Oxford University Press; 2014.
27. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int J Qual Methods*. 2006;5(1),80-92.
28. Nielsen J. *Usability Engineering*. San Francisco, CA: Morgan Kaufmann; 1994.
29. Khorsan R, Crawford C. How to assess the external validity and model validity of therapeutic trials: a conceptual approach to systematic review methodology. *Evid Based Complement Alternat Med*. 2014;2014,694804. doi:10.1155/2014/694804.
30. Schmuckler MA. What is ecological validity? A dimensional analysis. *Infancy*. 2001;2(4),419-436. doi:10.1207/S15327078IN0204_02.
31. Hartling L, Bond K, Santaguida PL, Viswanathan M, Dryden DM. Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. *J Clin Epidemiol*. 2011;64(8),861-871. doi:10.1016/j.jclinepi.2011.01.010.
32. Seo H-J, Kim SY, Lee YJ, et al. A newly developed tool for classifying study designs in systematic reviews of interventions and exposures showed substantial reliability and validity. *J Clin Epidemiol*. 2016;70,200-205. doi:<http://dx.doi.org/10.1016/j.jclinepi.2015.09.013>.
33. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8. doi:10.1186/1741-7015-8-24.
34. Mhaskar R, Djulbegovic B, Magazín A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol*. 2012;65(6),602-609. doi:10.1016/j.jclinepi.2011.10.016.
35. Sandelowski M, Barroso J. Reading qualitative studies. *Int J Qual Methods*. 2002;1(1),74-108. doi:10.1177/160940690200100107.

36. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: Wiley Online Library; 2008.
37. Faggion CM, Jr. Risk of bias assessment should not go beyond reporting assessment. *J Clin Epidemiol*. 2016;72,126-127. doi:10.1016/j.jclinepi.2015.11.014.
38. 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(10),1425-1434. doi:10.1177/1049732312452937.
39. Verhage A, Boels D. Critical appraisal of mixed methods research studies in a systematic scoping review on plural policing: assessing the impact of excluding inadequately reported studies by means of a sensitivity analysis. *Qual Quant*. 2016;51(4),1449-1468. doi:10.1007/s11135-016-0345-y.
40. Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychol Assess*. 1995;7(3),238.
41. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health*. 2010;25(10),1229-1245. doi:10.1080/08870440903194015.
42. Olsson C, Ringner A, Borglin G. Including systematic reviews in PhD programmes and candidatures in nursing - 'Hobson's choice'? *Nurse Educ Pract*. 2014;14(2),102-105. doi:10.1016/j.nepr.2014.01.005.