International Journal of Nursing Studies, 2012, 49(1):47-53. DOI: 10.1016/j.ijnurstu.2011.07.002

Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review

Romina Pace[a], Pierre Pluye[a], Gillian Bartlett[a], Ann C. Macaulay[b], Jon Salsberg [b], Justin Jagosh [b], Robbyn Seller [b]

a. Department of Family Medicine, McGill University, 517 Pine Avenue West, Montreal, QC, Canada H2W 1S4
b. Center for Participatory Research at McGill, Department of Family Medicine, McGill University, Canada

ABSTRACT

Background: Systematic literature reviews identify, select, appraise, and synthesize relevant literature on a particular topic. Typically, these reviews examine primary studies based on similar methods, e.g., experimental trials. In contrast, interest in a new form of review, known as mixed studies review (MSR), which includes qualitative, quantitative, and mixed methods studies, is growing. In MSRs, reviewers appraise studies that use different methods allowing them to obtain in-depth answers to complex research questions. However, appraising the quality of studies with different methods remains challenging. To facilitate systematic MSRs, a pilot Mixed Methods Appraisal Tool (MMAT) has been developed at McGill University (a checklist and a tutorial), which can be used to concurrently appraise the methodological quality of qualitative, quantitative, and mixed methods studies.

Objectives: The purpose of the present study is to test the reliability and efficiency of a pilot version of the MMAT.

Methods: The Center for Participatory Research at McGill conducted a systematic MSR on the benefits of Participatory Research (PR). Thirty-two PR evaluation studies were appraised by two independent reviewers using the pilot MMAT. Among these, 11 (34%) involved nurses as researchers or research partners. Appraisal time was measured to assess efficiency. Inter-rater reliability was assessed by calculating a kappa statistic based on dichotomized responses for each criterion. An appraisal score was determined for each study, which allowed the calculation of an overall intra-class correlation.

Results: On average, it took 14 min to appraise a study (excluding the initial reading of articles). Agreement between reviewers was moderate to perfect with regards to MMAT criteria, and substantial with respect to the overall quality score of appraised studies.

Conclusion: The MMAT is unique, thus the reliability of the pilot MMAT is promising, and encourages further development.
Keywords: Mixed methods research, Literature review, Systematic mixed studies review, Critical appraisal tool

WHAT IS ALREADY KNOWN ABOUT THIS TOPIC?

- Critical appraisal is an important stage in undertaking systematic literature reviews.
- Guidance exists for appraising the methodological quality of qualitative and quantitative studies included in systematic reviews.
- Mixed studies reviews include qualitative, quantitative, and mixed methods studies, and can provide greater understanding of a health issue.

WHAT THIS PAPER ADDS

- It reports the test of the efficiency and reliability of a unique Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies reviews.
- This tool allows the concomitant quality appraisal of qualitative, quantitative, and mixed methods studies.
- This test of the pilot MMAT is encouraging, and leads to the proposal of a 2011 version of the MMAT.

1. INTRODUCTION

Interest in the concomitant review of qualitative, quantitative, and mixed methods studies, known as a mixed studies review (MSR), is growing (Grant and Booth, 2009), particularly in health sciences (Pluye et al., 2009). MSRs address complex questions comprising qualitative and quantitative aspects. For example, in a MSR examining the question 'What are the impacts of clinical information retrieval technology?' types of impact were based on findings of qualitative studies, and then the importance of positive impacts was estimated using results of quantitative studies (Pluye et al., 2005). This new form of literature review has the potential to provide a rich, detailed, and highly practical understanding of complex health interventions and programs, which can be more relevant to and useful for clinicians and decision-makers. For example, ''examining the effectiveness of interventions to increase the uptake of breast-feeding [based on results of quantitative studies] benefits from examining reasons why people do and do not breastfeed, their perceptions of the advantages of not doing so, and obstacles to this practice [based on findings of qualitative research studies]'' (Sheldon, 2005, p. 5).

In MSRs, reviewers apply mixed methods research to review the literature. The foundation of mixed methods research is to combine the strengths of qualitative and quantitative methods by integrating the in-depth descriptions of complex phenomena obtained by qualitative methods with the statistical generalizability of quantitative methods. The conceptualization of mixed methods research is new and no standard valid critical appraisal tool for mixed methods research exists (Creswell and Plano Clark, 2007; O'Cathain et al., 2008; O'Cathain, 2010), whereas, multiple standard tools exist for quantitative methods, and a few valid tools exist for qualitative methods (Crowe and Sheppard, 2011; EQUATOR, 2011; Simera et al., 2010).

When conducting systematic MSRs, reviewers identify, select, appraise, and synthesize relevant qualitative, quantitative, and mixed methods studies, and as with all systematic reviews, the appraisal of the methodological quality of included studies is crucial. The content validation of an initial version of a critical appraisal tool for systematic MSRs, called the Mixed Methods Appraisal Tool (MMAT), has previously been reported in the International Journal of Nursing Studies (Pluye et al., 2009). The MMAT is unique in that no other appraisal tool for systematic MSRs considers all study designs, including mixed methods research designs (Crowe and Sheppard, 2011; Simera et al., 2010). The purpose of the present paper is to describe the reliability and efficiency of the pilot MMAT.

2. BACKGROUND

Pluye et al. (2009) reported a qualitative thematic data analysis of the quality appraisal procedures used in 17 systematic health-related MSRs to determine the criteria without which a judgment on quality cannot be made for qualitative, quantitative, and mixed methods studies. Based on this analysis, an initial 15-criteria MMAT was proposed. The purpose of this tool was to allow for the concurrent appraisal of studies employing the most common methodologies and methods, with a set of a few generic quality criteria.

The MMAT contains five specific sets of criteria: (1) a 'qualitative' set for qualitative studies, and qualitative components of mixed methods research; (2) a 'randomized controlled' set for randomized controlled quantitative studies, and randomized controlled components of mixed methods research; (3) a 'non-randomized' set for non- randomized quantitative studies, and non-randomized components of mixed methods research, (4) an 'observational descriptive' set for observational descriptive quantitative studies, and observational descriptive components of mixed methods research; and (5) a set 'mixed methods' for mixed methods research studies. Each study type is judged within its methodological domain. For example, appraising the quality of a cohort study involves the 'non- randomized' set.

Furthermore, appraising a mixed methods study involves three sets: the 'qualitative' set, the

appropriate quantitative set, and the 'mixed methods' set. For instance, the most frequent mixed methods studies combine qualitative research, e.g., an ethnographic study, and a descriptive observational quantitative study, e.g., a cross-sectional prevalence survey (Bryman, 2006). Appraising such combinations using the MMAT involves three sets: (1) the 'qualitative' set for appraising the ethnographic aspect of the study, (2) the 'observational descriptive' set for the quantitative survey aspect, and (3) the 'mixed methods' set for appraising the integration between the qualitative and the quantitative aspects.

Alone, the 'quality score' derived from the MMAT is not very informative to report a critical appraisal. Describing the quality of studies using MMAT criteria is more informative. The MMAT does not position qualitative, randomized controlled, non-randomized, observational descriptive, and mixed methods studies in a hierarchy of evidence, but the 'quality score' might offer a rationale for excluding primary studies of low quality within each methodological domain, e.g., no criteria met (score = 0). For mixed methods studies, we argue that the overall quality of the combination of qualitative and quantitative methods cannot exceed the quality of its weakest component. As such, the overall quality score is the lowest score of the study components (qualitative, or quantitative, or mixed).

3. METHODS

The Center for Participatory Research at McGill (PRAM) conducted a review on the benefits of participatory research (PR) in the health sciences. PR is a collaborative approach to research involving both researchers and those affected by the research throughout the research process (Macaulay et al., 1999). Given the heterogeneity of methods used across PR projects, this review presented an opportunity to test the MMAT.

3.1. Development of a pilot tool

During the summer of 2009, the MMAT was tested using six PR studies, and the initial MMAT criteria were revised. Consensus on the revised criteria was reached through discussions among four reviewers, which led to some improvement and a 19-criteria pilot MMAT. For example, there was not enough information to apply the 'randomized controlled' criterion ''complete outcome data and/or low withdrawal/drop-out'', and the criterion was changed to ''complete outcome data (80% or above) and low withdrawal/drop-out (below 20%)''. These additional limits were in line with accepted values (Phillips et al., 2009). Furthermore, the 'non-randomized' set of criteria was added (adapted from Wells et al., 2009).

3.2. Testing efficiency and reliability

The pilot MMAT contained an appraisal form and a tutorial. The tutorial included definitions and examples to aid reviewers. In January 2010, 23 PR programs had been identified for the PRAM review. From this sample, four were excluded because there were no evaluation studies. Among the retained 19 programs, some included more than one evaluation study and, as such, 32 evaluation studies were included (11 involving nurses as PR researchers or partners). One of these was a mixed methods study with three components (a qualitative, an observational and a mixed methods component). Therefore, 34 study components were independently appraised by two reviewers with the MMAT (Table 1).

With regards to testing the efficiency of the MMAT, the time needed to read studies for inclusion/exclusion was not counted since this does not depend on the MMAT, but is associated with the length and complexity of both publications and studies. Only the time required to re-read articles for critical appraisal was recorded. A mean time was calculated from the two reviewers' times.

In keeping with Carmines and Zeller (1979), we defined reliability as the extent to which an assessment provides the same results in different situations; for example when the quality appraisal of one study is conducted by different reviewers. For each criterion, the presence or absence was reported as 1 and 0, respectively. Then, the reviewers discussed their responses. For the dichotomized responses pre- and post-discussion, SPSS 18 software was used to calculate the kappa statistic (Landis and Koch, 1977). A negative kappa was interpreted as indicating no agreement; a kappa between 0 and 0.20, slight agreement; between 0.21 and 0.40, fair agreement; between 0.41 and 0.60, moderate agreement; between 0.61 and 0.80, substantial agreement; and between 0.81 and 1.00, almost perfect agreement (Landis and Koch, 1977).

Subsequently, a 'quality score' for each evaluation study was calculated as a percentage: (number of 'yes' responses divided by the number of 'applicable criteria') x 100 (e.g., a study with all qualitative and quantitative observational components present, along with good overall mixed methods approach would be scored as 100%: [(6 + 3 + 3)/ 12] x 100). To examine the inter-rater reliability for the total appraisal score (continuous data), an intra-class correlation (ICC) was calculated (Shrout and Fleiss, 1979). SPSS 18 software was used to calculate the ICC using a two-way mixed model (absolute agreement type), and the ICC was interpreted similar to kappa (Garson, 2010).

4. RESULTS

On average, it took approximately 14 min to appraise a study (range: 4–40min). The consistency of the global 'quality score' between reviewers (ICC) was 0.72 pre- and 0.94 post-discussion (Table 2).

Inter-rater reliability pre-discussion: With respect to 17 of the 19 criteria, there was almost perfect agreement for 7 criteria, substantial agreement for 1 criterion, moderate agreement for 3 criteria,

fair agreement for 4 criteria, slight agreement for 1 criterion, and no agreement for only 1 criterion (Table 3). Inter-rater reliability post-discussion: There was almost perfect agreement for 13 criteria, substantial agreement for 2 criteria, and moderate agreement for 2 criteria (Table 3). With regards to the two remaining criteria (1.1 and 3.3), one or both reviewers gave a consistent score for all studies, precluding the calculation of a kappa. However, for these two criteria, inter-rater agreement was 88.9% and 83.3%, respectively.

5. DISCUSSION

Results suggest the pilot MMAT was easy to use. Inter-rater reliability scores ranged from moderately reproducible to perfect agreement. After discussion, the raters were able to reach a consensus on 19 (76%) of the 25 pre- discussion disagreements. These disagreements were, for the most part, resolved by referring to the MMAT tutorial. The sets of criteria with the most discordant results pre-discussion were the 'non-randomized' (32%) and the 'qualitative' (48%) sets. These differences may be interpreted as follows: (1) evaluative criteria for the quality of non-randomized studies are newly established (Wong et al., 2008), and (2) the use of critical appraisal tools to evaluate qualitative research is complex (Cohen and Crabtree, 2008).

Critically appraising qualitative research studies remains controversial (Cohen and Crabtree, 2008; Murphy et al., 1998; Pope and Mays, 2009; Pope et al., 2007; Sandelowski et al., 2007). Main issues are the diversity of qualitative researchers' worldviews (Niglas, 2010), and the different characteristics of qualitative research methods. For instance, the Cochrane Qualitative Research Methods Group proposes eight tools for critically appraising qualitative research, but does not specifically recommend the usage of tools with specific or generic criteria: ''A range of appraisal instruments and frameworks is available for use in the assessment of the quality of qualitative research. Some are generic, being applicable to almost all qualitative research designs; others have specifically been developed for use with certain methods or techniques'' (CQRMG, 2010). Firstly, some critical appraisal tools include specific criteria aligned with established qualitative traditions, e.g., ethnography, and criteria for qualitative research designs that do not fit with these traditions, e.g., interpretive description. In fact, the methodology of qualitative research studies in the health sciences is often inappropriately labeled as biography, case study, ethnography, grounded theory, or phenomenology, when it would be more appropriate to conceive it as qualitative or interpretive description (Caelli et al., 2003; Sandelowski, 2010; Thorne et al., 1997). Secondly, other critical appraisal tools use generic criteria such as the CASP tool that includes 10 questions to assist health professionals to specifically evaluate the rigor, credibility and relevance to practice of qualitative studies (Critical Appraisal Skills Program, 2011). However, the

comparison of three online critical appraisal tools for qualitative research, including the CASP tool, concluded that more evaluation is needed to examine whether a few generic criteria are likely to be appropriate and representative for the most common types of qualitative research (Hannes et al., 2010).

The 'qualitative' set of the MMAT is based on few generic criteria since the purpose is to describe and compare similar characteristics of studies within each methodological domain. Therefore, the MMAT must be used with caution, and further content validation and reliability testing involving multiple reviewers and a larger sample of studies is needed. Indeed, developing tools with measurement properties is an iterative process (Vogt et al., 2004), and these encouraging pilot results call for further steps. This is not trivial as the MMAT is unique. O'Cathain (2010) reviewed the literature on appraisal tools for mixed methods research, and found no tools for systematic MSRs outside the MMAT. Crowe and Sheppard (2011) reviewed the literature on all types of appraisal tools, and found six tools for all research designs, and the MMAT, was the only one of these tools to appraise mixed methods research. Notably, Crowe and Sheppard (2011) mentioned that the reliability of the MMAT was unknown. The present pilot work contributes to address this issue.

6. CONCLUSION

Our results suggest the MMAT is promising. Reliability is a key property of a critical appraisal tool, and the efficiency is important from a reviewer's perspective. In 2010, the pilot MMAT was used and discussed in four 90-min workshops that suggested further refinement of criteria. These workshops involved diverse audiences such as graduate students enrolled in a mixed methods research course, researchers and research professionals with experience in qualitative, quantitative, and mixed methods research, and members of the Cochrane collaboration with experience in systematic MSRs or in systematic reviews of qualitative studies, e.g., meta-ethnography, or of quantitative studies, e.g., systematic reviews of randomized controlled trials. The MMAT has been revised using feedback from these workshops and also the first published comprehensive framework for assessing the quality of mixed methods research (O'Cathain, 2010). The 2011 version of the MMAT checklist is presented in Appendix A.

ACKNOWLEDGEMENTS

Romina Pace holds a Summer Research Bursary from the Faculty of Medicine, McGill University. Pierre Pluye holds a New Investigator Award from the Canadian Institutes of Health Research (CIHR). The present work is supported by CIHR and the Center for Participatory Research at McGill (PRAM).

CONTRIBUTIONS

Pierre Pluye, Marie-Pierre Gagnon, Frances Griffiths, and Janique Johnson-Lafleur proposed an initial version of MMAT criteria. Romina Pace and Pierre Pluye led the test of the pilot MMAT (criteria and tutorial). Gillian Bartlett, Belinda Nicolau, Robbyn Seller, Justin Jagosh, Jon Salsberg, and Ann C. Macaulay contributed to this test. Émilie Robert, Margaret Cargo, Alicia O'Cathain, Frances Griffiths, Felicity Boardman, Marie-Pierre Gagnon, and Gillian Bartlett contributed to the 2011 version of the MMAT (Appendix A).

CONFLICT OF INTEREST STATEMENT

None declared.

FUNDING

Canadian Institutes of Health Research (CIHR): Funding; Center for Participatory Research at McGill (PRAM): Funding, co-investigation, and co-authorship.

ETHICAL APPROVAL

Not applicable.

REFERENCES

Bryman, A., 2006. Integrating quantitative and qualitative research: how is it done? Qualitative Research 6 (1), 97–113.

Caelli, K., Ray, L., Mill, J., 2003. 'Clear as Mud': toward greater clarity in generic qualitative research. International Journal of Qualitative Methods 2 (2), 1–23.

Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment. Sage, Thousand Oaks.

Cohen, D.J., Crabtree, B.F., 2008. Evaluative criteria for qualitative research in health care: controversies and recommendations. Annals of Family Medicine 6 (4), 331–339.

Cochrane Qualitative Research Methods Group (CQRMG), 2010. Critical Appraisal of Qualitative Research (Draft Chapter – in peer review with Cochrane Handbook Editors). (retrieved August 27.08.10)http://www.joannabriggs.edu.au/cqrmg/documents/Cochrane_Guidance/Chapter6_Guidance_Critical_Appraisal.pdf.

Creswell, J., Plano Clark, V.L., 2007. Designing and Conducting Mixed Methods Research. Sage, Thousand Oaks.

Critical Appraisal Skills Program (CASP), 2011. Ten Questions to Help You Make Sense of Qualitative Research. (retrieved 15.01.11)www.sph.nhs.uk/what-we-do/public-health-workforce/ resources.

Crowe, M., Sheppard, L., 2011. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. Journal of Clinical Epidemiology 64 (1), 79–89.

EQUATOR, 2011. The EQUATOR Network Website: The Resource Centre for Good Reporting of Health Research Studies. (retrieved 30.01.11)www.equator-network.org/resource-centre.

Garson, D.G., 2010. Reliability Analysis: Statnotes, from North Carolina State University, Public Administration Program., Available from: http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm#rater.

Grant, M.J., Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal 26 (2), 91–108.

Hannes, K., Lockwood, C., Pearson, A., 2010. A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. Qualitative Health Research 20 (12), 1736–1743.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Macaulay, A.C., et al., 1999. Participatory research maximises community and lay involvement. British Medical Journal 319 (7212), 774–778.

Murphy, E., Dingwall, R., Greatbatch, D., Parker, S., Watson, P., 1998. Qualitative research methods in health technology assessment: a review of the literature. Health Technology Assessment 2 (16), 1–274.

Niglas, K., 2010. The multidimensional model of research methodology: an integrated set of continua. In: Tashakkori, A., Teddlie, C. (Eds.), Handbook of Mixed Methods in Social and Behavioral Research. Sage, Thousand Oaks, pp. 215–236.

O'Cathain, A., Murphy, E., Nicholl, J., 2008. The quality of mixed methods studies in health services research. Journal of Health Services Research & Policy 13 (2), 92–98.

O'Cathain, A., 2010. Assessing the quality of mixed methods research: towards a comprehensive framework. In: Tashakkori, A., Teddlie, C. (Eds.), Handbook of Mixed Methods in Social and Behavioral Research. Sage, Thousand Oaks, pp. 531–555.

Phillips, B., Ball, C., Sackett, D., Badenoch, D., Straus, S., Haynes, B., et al., 2009. Oxford Centre for Evidence-based Medicine: Levels of Evidence. , Available from: http://www.cebm.net/?o=1025.

Pluye, P., Grad, R.M., Dunikowski, L., Stephenson, R., 2005. Impact of clinical information-retrieval technology on physicians: a literature review of quantitative, qualitative and mixed methods

studies. International Journal of Medical Informatics 74 (9), 745–768.

Pluye, P., Gagnon, M.P., Griffiths, F., Johnson-Lafleur, J., 2009. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. International Journal of Nursing Studies 46 (4), 529–546.

Pope, C., Mays, N., 2009. Critical reflections on the rise of qualitative research. British Medical Journal 339, b3425.

Pope, C., Mays, N., Popay, J., 2007. Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods. Open University Press, Berkshire.

Sandelowski, M., 2010. What's in a name? Qualitative description revisited. Research in Nursing and Health 33 (1), 77–84.

Sandelowski, M., Voils, C.I., Barroso, J., 2007. Comparability work and the management of difference in research synthesis studies. Social Science & Medicine 64 (1), 236–247.

Sheldon, T.A., 2005. Making evidence synthesis more useful for management and policy-making. Journal of Health Services Research and Policy 10 (3 Suppl. 1), 1–5.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86 (2), 420–428.

Simera, I., Moher, D., Hoey, J., Schulz, K.F., Altman, D.G., 2010. A catalogue of reporting guidelines for health research. European Journal of Clinical Investigation 40 (1), 35–53.

Thorne, S., Kirkham, S.R., MacDonald-Emes, J., 1997. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. Research in Nursing and Health 20 (2), 169–177.

Vogt, D.S., King, D.W., King, L.A., 2004. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychological Assessment 16 (3), 231–243.

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., et al., 2009. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-analyses. , Available from: http:// www.ohri.ca/programs/clinical_epidemiology/oxford.htm.

Wong, W.C., Cheung, C.S., Hart, G.J., 2008. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerging Themes in Epidemiology 5, 23, doi:10.1186/1742-7622-5-23.

TABLES AND FIGURES

Table 1 Study components appraised for reliability test.

Study components	Number[a]
1. Qualitative	9
2. Randomized controlled	8
3. Non-randomized controlled	6
4. Observational (no control group)	10
5. Mixed methods	1

[a] In total, 32 studies were appraised, including one mixed methods study with a qualitative component, an observational component, and a mixed methods component. Thus, the sum of components in the second column is 34.

Table 2 Intra-class correlation for global appraisal score.

	ICC (95% CI)
Pre-discussion	0.717 (0.485–0.853)
Post-discussion	0.936 (0.872–0.968)

ICC: intra-class correlation; CI: confidence interval

Table 3 Kappa scores per criterion.

Criteria by type of study (or study component of mixed methods research)	Pre-discussion		Post-discussion	
	Kappa	*p*-Value	Kappa	*p*-Value
Qualitative				
1.1 Qualitative objective or question	NA	NA	NA	NA
1.2 Appropriate qualitative approach or method	0.526	0.073	1	0.003
1.3 Description of the context	1	NA	1	NNA
1.4 Description of participants and sampling	0.250	0.257	0.526	0.073
1.5 Description of data collection and analysis	-0.174	0.571	1	0.003
1.6 Discussion of researchers' reflexivity	0.400	0.134	1	0.003
Randomized controlled				
2.1 Appropriate sequence generation/randomization	1	0.005	1	0.005
2.2 Allocation concealment and/or blinding	1	0.005	1	0.005
2.3 Complete outcome data and low withdrawal	1	0.005	1	0.005
Non-randomized				
3.1 Recruitment in a way that minimized confounders	0.333	0.273	0.571	0.121
3.2 Intervention and control group comparable	0.333	0.414	1	0.014
3.3 Evidence of an absence of contamination	NA	NA	NA	NA
3.4 Complete outcome date/acceptable response rate	0	1	1	0.014
Observational descriptive				
4.1 Appropriate sampling and sample	0.615	0.035	1	0.005
4.2 Justification of measurement (valid/standard)	0.545	0.053	0.714	0.035
4.3 Acceptable response rate	0.600	0.038	0.783	0.011
Mixed methods				
5.1 Combination of qualitative and quantitative data collection-analysis techniques or procedures	1	NA	1	NA
5.2 Justification of the mixed methods design	1	NA	1	NA
5.3 Integration of qualitative and quantitative data or results	1	NA	1	NA

NA: not applicable.

APPENDIX

Appendix A. The 2011 MMAT checklist

Types of mixed methods study components or primary studies	Methodological quality criteria (see tutorial for definitions and examples)	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	• Are there clear qualitative and quantitative research questions (or objectives), or a clear mixed methods question (or objective)?[a] • Do the collected data allow address the research question (objective)? E.g., consider whether the follow-up period is long enough for the outcome to occur (for longitudinal studies or study components). • Further quality appraisal may be not feasible when the answer is 'No' or 'Can't tell' to one or both screening questions.				
1. Qualitative	1.1. Are the sources of qualitative data (archives, documents, informants, observations) relevant to address the research question (objective)? 1.2. Is the process for analyzing qualitative data relevant to address the research question (objective)?				

	1.3. Is appropriate consideration given to how findings relate to the context, e.g., the setting, in which the data were collected? 1.4. Is appropriate consideration given to how findings relate to researchers' influence, e.g., through their interactions with participants?
2. Quantitative randomized controlled (trials)	2.1. Is there a clear description of the randomization (or an appropriate sequence generation)? 2.2. Is there a clear description of the allocation concealment (or blinding when applicable)? 2.3. Are there complete outcome data (80% or above)? 2.4. Is there low withdrawal/drop-out (below 20%)?
3. Quantitative non-randomized	3.1. Are participants (organizations) recruited in a way that minimized selection bias? 3.2. Are measurements appropriate (clear origin, or validity known, or standard instrument; and absence of contamination between groups when appropriate) regarding the exposure/intervention and outcomes? 3.3. In the groups being compared (exposed vs. non-exposed; with intervention vs. without; cases vs. controls), are the participants comparable, or do researchers take into account (control for) the difference between these groups? 3.4. Are there complete outcome data (80% or above), and, when applicable,

	an acceptable response rate (60% or above), or an acceptable follow-up rate for cohort studies (depending on the duration of follow-up)?
4. Quantitative descriptive	4.1. Is the sampling strategy relevant to address the quantitative research question (quantitative aspect of the mixed methods question)? 4.2. Is the sample representative of the population understudy? 4.3. Are measurements appropriate (clear origin, or validity known, or standard instrument)? 4.4. Is there an acceptable response rate (60% or above)?
5. Mixed methods	5.1. Is the mixed methods research design relevant to address the qualitative and quantitative research questions (or objectives), or the qualitative and quantitative aspects of the mixed methods question (or objective)? 5.2. Is the integration of qualitative and quantitative data (or results) relevant to address the research question (objective)?a 5.3. Is appropriate consideration given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results) in a triangulation design?
	Criteria for the qualitative component (1.1 to 1.4), and appropriate criteria for the quantitative component (2.1 to 2.4, or 3.1 to 3.4, or 4.1 to 4.4), must be also applied.

Note: The 2011 MMAT is comprised of two parts: the above checklist and a tutorial. The tutorial is available on the following free public wiki website: http://

mixedmethodsappraisaltoolpublic.pbworks.com. Please contact pierre.pluye@mcgill.ca for dissemination, application, and feedback.

[a] This item is not considered as double-barreled question since in mixed methods research, qualitative and quantitative data may be integrated, and/or qualitative findings and quantitative results can be integrated.