

Creation of an online inventory for choosing critical appraisal tools

Quan Nha Hong^{a,*}, Julien Bouix-Picasso^{b,c} and Christian Ruchon^d

^a*School of Rehabilitation, Université de Montréal, Montreal, QC, Canada*

^b*Faculty of Nursing Sciences, Université de Montréal, Montreal, QC, Canada*

^c*Laboratoire Éducatifs et Promotion en Santé, Université Sorbonne Paris Nord, Bobigny, France*

^d*Department of Family Medicine, McGill University, Montreal, QC, Canada*

Critical appraisal of evidence is performed to assess its validity, trustworthiness and usefulness in evidence-based practice. There currently exists a large number and variety of critical appraisal tools (also named risk of bias tools and quality assessment instruments), which makes it challenging to identify and choose an appropriate tool to use. We sought to develop an online inventory to inform librarians, practitioners, graduate students, and researchers about critical appraisal tools. This online inventory was developed from a literature review on critical appraisal tools and is kept up to date using a crowdsourcing collaborative web tool (eSRAP-DIY). To date, 40 tools have been added to the inventory (www.catevaluation.ca), and grouped according to five general categories: (a) quantitative studies, (b) qualitative studies, (c) mixed methods studies, (d) systematic reviews and (e) others. For each tool, a summary is provided with the following information: tool name, study designs, number of items, rating scale, validity, reliability, other information (such as existing websites or previous versions), and main references. Further studies are needed to test and improve the usability of the online inventory, and to find solutions to reduce to monitoring and update workload.

Keywords: Risk of bias, critical appraisal tool, quality appraisal, quality assessment, inventory

1. Introduction

The purpose of this article is to describe the development and maintenance of an online inventory of critical appraisal tools (CATs). CATs (also named risk of bias tools and quality assessment instruments) are checklists or scales that provide a list of domains and items to assess the quality of studies (Bai et al., 2012). These tools can be used as part of evidence-based practice that advocates to critically appraise the evidence to assess its validity, trustworthiness, and usefulness (Burls, 2009; Greenhalgh, 2014). Also, critical appraisal is a step of systematic literature reviews to ensure that conclusions drawn from the included studies properly reflect the quality of evidence reviewed. The appraisal can help to highlight the strengths and weaknesses of the evidence reviewed, to select studies, to indicate how much confidence readers can have in the results, to guide the synthesis and interpretation of the findings, and to identify the gaps in knowledge (Booth et al., 2016; Viswanathan et al., 2012).

*Corresponding author: Quan Nha Hong, School of Rehabilitation, Université de Montréal, Montreal, QC, Canada. E-mail: quan.nha.hong@umontreal.ca.

A variety of CATs has been developed to formalize the appraisal process and ensure it is done in a more rigorous, systematic, and explicit manner. Despite the numerous existing literature reviews analyzing over 500 CATs, there is still no clear guidance to help select which of the many and varied CATs to use. These reviews covered different study designs (e.g., randomized controlled trials (RCT) (Armijo-Olivo et al., 2013; Moher et al., 1995), non-randomized studies (Deeks et al., 2003; Jarde et al., 2012; Saunders et al., 2003; Shamliyan et al., 2010), or qualitative studies (Majid & Vanstone, 2018; Santiago-Delefosse et al., 2016; Walsh & Downe, 2006) and analyzed the tools using different criteria (e.g., number of items, time needed to apply, content, frequency of use). Also, several tools were not developed using methodological standards or are not supported by sound validation and reliability testing (Crowe & Sheppard, 2011; Katrak et al., 2004). This makes it more challenging for librarians, practitioners, graduate students and researchers to identify and choose a proper tool to use. To respond to this challenge, we developed an online inventory of CATs.

2. Online inventory of critical appraisal tools

The online public inventory of CATs, <https://www.catevaluation.ca/>, was launched in May 2020. We built the inventory in WordPress, working with a web developer experienced in using this software. Its development was inspired by existing websites such as the EQUATOR Network on reporting guidelines (Altman et al., 2008). Users can consult the inventory in either English or French, search for a specific CAT using the search bar, and using the study design algorithm on the home page, they can find CATs for a specific study design or category of studies.

The initial inventory in May 2020 consisted of 37 CATs. These CATs were identified from the results of literature reviews performed during the doctoral project of the first author (Hong, 2018). To select the CATs for the inventory, the selection criteria listed in Table 1 were used. Given that tools relying on rigorous development and testing are being advocated for in the literature, we decided to include only tools that were supported by validation and/or reliability testing (Whiting et al., 2017). When several tools for a specific study design met the selection criteria, we decided to only include the most recent ones. In the online inventory, these CATs are grouped into five general categories: (a) quantitative studies, (b) qualitative studies, (c) mixed methods studies, (d) systematic reviews, and (e) others. Each tool is described in a table that presents the tool name, study design, number of items, rating scale, validity, reliability, other information (such as existing websites or previous versions), and main references.

To keep the inventory up to date, we are monitoring the CAT literature using the collaborative web tool eSRAP©-DIY, <https://esrap-spor-diy.ca>. eSRAP-DIY was developed to monitor and filter the latest most relevant scholarly publications on a given topic (Granikov et al., 2020). Using a pre-defined search strategy, this system

Table 1
Selection criteria

Criteria	Inclusion	Exclusion
Tool	List of criteria to judge the methodological quality of studies.	Grading approaches.
Quality dimension	Tools including methodological quality criteria.	Tools limited to reporting quality criteria.
Type of tool	Tools for primary studies and systematic reviews in health research.	Tools for guidelines, animal studies, and <i>in vitro</i> studies. Tools limited to external validity.
Validity and/reliability (at least one of these criteria)	<ul style="list-style-type: none"> – The tool development included consultations with experts (e.g., Delphi study, survey). – The tool was compared with other existing tools or expert judgment. – The tool was pilot tested with experts/users, and results were used to refine the tool. – Factor analysis was performed. – Correlations with related or unrelated constructs were calculated. – Correlations between the items of the tool were performed (internal consistency). – Ratings from two or more reviewers that appraised the same studies with the tool were compared (inter-rater reliability). – Ratings from reviewers that rated the same studies twice at different times were compared (test-retest reliability). 	No information provided on validity or reliability.
Language	English or French.	Other languages than English or French.

monitors one or several bibliographic databases and retrieves new records (titles and abstracts) as soon as they are added. Then, through crowdsourcing, the retrieved records results are screened using one simple question based on the selection criteria in Table 1: *Is this paper about a critical appraisal tool/risk of bias tool and presents information on its development and/or validity and/or reliability?*. This accelerates the screening process and reduces the workload of each member of the ‘crowd’. A search strategy to identify papers on CATs was developed for Scopus and added to the eSRAP-DIY system in December 2019. The search is run weekly and includes terms related to quality assessment, validity, reliability, and tools. Since May 2020, a crowd of three has been screening the records retrieved by eSRAP-DIY. Between May 2020 and November 2021, 847 records were identified in eSRAP-DIY and, among these, the crowd identified three new tools and 13 new studies about tools already in the inventory. To date, a total of 40 CATs has been added in the online inventory.

3. Discussion

For each CAT user, finding and choosing a proper CAT to assess the quality of

a study can be a daunting task because of the large number and variety of existing tools. To address this issue, we developed an online inventory to inform librarians, practitioners, graduate students and researchers about the CATs that have been tested for their validity and/or reliability. We structured the inventory into five general research categories to facilitate the identification of a tool, and provided a short summary of each tool. This categorization could help users identify and select CATs by providing them a quick overview of the available validated tools.

Maintaining online inventories up to date and relevant for users is a challenge. The success of the maintenance of our CAT inventory currently relies on voluntary efforts of people interested in contributing to a collective effort to screen records identified in the monitoring system eSRAP-DIY. In addition to the first author, five raters were recruited to participate to the screening of records in eSRAP-DIY. However, only two logged in the systems and rated records. Participant retention is a problem that is often observed when using crowdsourcing (Strong & Simmons, 2018). Also, compared to systematic reviews that typically have an end-point, monitoring records for an online inventory is an ongoing process, and thus, it is more challenging to keep the crowd motivated. Currently, the workload to screen new records remains manageable (around 600 new records per year), but is likely to increase yearly similarly to the annual growth in the number of articles published, which is estimated to be approximately 3% (Ware & Mabe, 2015). In order to improve the engagement of the crowd, it is necessary to provide frequent feedback on the crowd's performance (Strong & Simmons, 2018) such as sending a newsletter email to the raters on a regular basis to inform the progress of the crowd or notifications of new records.

There are some limitations of this inventory. First, it is possible that not all eligible CATs were identified and included in the inventory. We limited our search to articles indexed in bibliographic databases and did not search for tools published in the grey literature. Also, eSRAP-DIY is only monitoring one bibliographic database (SCOPUS). Second, we adapted the selection criteria depending on the tool category. For example, on the one hand, given that several CATs have been developed for the RCTs and non-randomized studies categories, we mainly retained the most recently developed tools that had information on both validity and reliability. On the other hand, given the paucity of existing tools for the mixed methods and systematic reviews categories, our inclusion criteria were less stringent; we retained tools even if there was information only on validity or only on reliability. Third, we selected the CATs based on whether they included validity and/or reliability testing, regardless of the validity and/or reliability results. We advise the users of our inventory to read the results of the studies listed in the CAT descriptive tables to make a decision regarding decide which tools to use.

4. Conclusion

This article presents the development and maintenance of an online inventory of CATs to help librarians, practitioners, graduate students and researchers to identify

and select the proper tools to use. Usability studies are necessary to test the ease of use, acceptability, and usefulness of our CAT inventory. There is also a need to explore other options to manual monitoring, such as machine learning approaches, that could help reduce the workload by reducing the number of irrelevant records to screen (O'Mara-Eves et al., 2015). We invite the research community to collaborate by informing us about new or missing CATs, providing feedback on the utility and usability, or joining the 'crowd' to screen records identified in the eSRAP-DIY system.

Acknowledgments

This project was funded by the Quebec Strategy for Patient-Oriented Research (SPOR) SUPPORT Unit. We would like to thank Robin Melanson for the website development, Vera Granikov and David Tang for their support with eSRAP-DIY, and Paula Bush for commenting an earlier version of this manuscript.

References

- Altman, D.G., Simera, I., Hoey, J., Moher, D., & Schulz, K. (2008). EQUATOR: Reporting guidelines for health research. *Open Medicine*, 2(2), e49.
- Armijo-Olivo, S., Fuentes, J., Ospina, M., Saltaji, H., & Hartling, L. (2013). Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. *BMC Medical Research Methodology*, 13(116), 1-19.
- Bai, A., Shukla, V.K., Bak, G., & Wells, G. (2012). *Quality assessment tools project report*. Canadian Agency for Drugs and Technologies in Health.
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE Publications.
- Burls, A. (2009). *What is critical appraisal?* (2nd ed.). Hayward Medical Communications.
- 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.
- Deeks, J.J., Dinnes, J., D'Amico, R., Sowden, A.J., Sakarovitch, C., Song, F., Petticrew, M., & Altman, D.G. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7(27), i-186.
- Granikov, V., Bouthillier, F., & Pluye, P. (2020). Understanding collaboration in monitoring research publications: Protocol for a qualitative multiple case study. *Education for Information*, 36(1), 69-79.
- Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. John Wiley & Sons.
- Hong, Q.N. (2018). Revision of the Mixed Methods Appraisal Tool (MMAT): A mixed methods study [Doctoral dissertation, McGill University]. Montréal, Canada.
- Jarde, A., Losilla, J.-M., & Vives, J. (2012). Methodological quality assessment tools of non-experimental studies: A systematic review. *Anales de Psicología*, 28(2), 617-628.
- Katrak, P., Bialocerkowski, A.E., Massy-Westropp, N., Kumar, S., & Grimmer, K.A. (2004). A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*, 4(22), 1-11.
- Majid, U., & Vanstone, M. (2018). Appraising qualitative research for evidence syntheses: A compendium of quality appraisal tools. *Qual Health Res*, 28(13), 2115-2131.
- Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16(1), 62-73.

- O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, 4(1), 5.
- Santiago-Delefosse, M., Gavin, A., Bruchez, C., Roux, P., & Stephen, S. (2016). Quality of qualitative research in the health sciences: Analysis of the common criteria present in 58 assessment guidelines by expert users. *Social Science & Medicine*, 148(1), 142-151.
- Saunders, L.D., Soomro, G.M., Buckingham, J., Jamtvedt, G., & Raina, P. (2003). Assessing the methodological quality of nonrandomized intervention studies. *Western Journal of Nursing Research*, 25(2), 223-237.
- Shamliyan, T., Kane, R.L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, 63(10), 1061-1070.
- Strong, L., & Simmons, R.K. (2018). *Citizen science: crowdsourcing for systematic reviews*. The Health Foundation, University of Cambridge.
- Viswanathan, M., Ansari, M.T., Berkman, N.D., Chang, S., Hartling, L., McPheeters, M., Santaguida, P.L., Shamliyan, T., Singh, K., Tsertsvadze, A., & Treadwell, J.R. (2012). *Assessing the risk of bias of individual studies in systematic reviews of health care interventions*. Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Comparative Effectiveness Reviews.
- Walsh, D., & Downe, S. (2006). Appraising the quality of qualitative research. *Midwifery*, 22(2), 108-119.
- Ware, M., & Mabe, M. (2015). *An overview of scientific and scholarly journal publishing*. STM – International Association of Scientific, Technical and Medical Publishers.
- Whiting, P., Wolff, R., Mallett, S., Simera, I., & Savović, J. (2017). A proposed framework for developing quality assessment tools. *Systematic Reviews*, 6(204), 1-9.