Analysis of Tools for the Critical Appraisal of Educational Research

Lo Sing Kai, Xie Qun, Ji Mingxia, So Winnie Wing Mui, *Hong Kong Institute of Education, Hong Kong*

Introduction

Employing inappropriate research methods or lacking explicitness in reporting are not suitable for guiding educational decisions or policy-making [1]. Educational researchers have started to pay more attention to study quality, which dictates the level of evidence generated [2]. With the rapid increase in the number of publications of journal articles, it is becoming more difficult to distinguish high-quality studies from those of lower quality [3]. Systematic reviews can overcome the above limitations; and can aggregate results to produce findings with greater power and reliability [2].

Although tools for methodological quality appraisal provide a means to identify the most rigorous research [1], each appraisal tool has been developed to fit specific objectives or designed with different criteria and philosophies in mind [4]. There has been little analysis on the different appraisal tools used in educational research. This study aims to systematically analyze appraisal tools used in systematic reviews of educational research. Specific objectives include an account of the appraisal tools for reviewing educational research, the nature of these tools, and the identification of common features of these tools.

Methods

Search and Inclusion of Critical Appraisal Tools

Databases searched included EBSCO, ProQuest, the British Education Index and the Australian Education Index. The keywords "systematic review" or "critical appraisal" and "education" (including MESH heading) were used to screen journal articles. In addition, a website search using the key phrase "systematic review critical appraisal education" was carried out to locate systematic review reports that might have not been found in the database search. All tools, checklists, rating scales or procedures retrieved in these two searching processes were kept at this stage and subjected to inclusion checking in the stage described below. For journal articles, all the searches were from inception to October 2013.

The tool had to be used in recently published journal articles, available online and ready to use [5]. The tool must also have been used at least twice of which at least once in the education field and the other time(s) can be in any field. If different versions of the same appraisal tool were used in different systematic reviews, the latest version would be included and analyzed.

Analysis of Appraisal Tools

Items in the tools were analysed using a 2-stage approach. First, each item was mapped into one of the four domains of quality assessment in the framework outlined in Furlong and Oancea [6]. The four domains are: (i) Epistemic Domain; (ii) Technological Domain; (iii) Capacity Development and Value for People Domain; and (iv) Economic Domain. Detailed analysis of the items was further carried out by mapping them into one of the sub-domains inside the four Domains. Then, items in the Epistemic Domain were further classified into six broad categories: (i) Study Design; (ii) Data Collection; (iii) Intervention; (iv) Data Analysis; and (v) Findings and Conclusions; and (vi) Overall Quality. Two independent reviewers carried out the coding. Any inconsistency in the coding was resolved through discussions between themselves or, when

necessary, with a senior author who has a track record in critical appraisal and systematic reviews.

Results

Basic characteristics of the appraisal tools

A total of 12 appraisal tools were included in the analysis based on the prescribed selection criteria. Seven tools were originally developed in medical or social sciences areas and applied to educational research subsequently; and five tools were originally developed for disciplinary areas in education. Most appraisal tools were designed for studies employing quantitative (8 out of 12) or mixed (3 out of 12) research methods; only one tool (JBI-QARI; Australia; 2007) was designed for qualitative study. Table 1 describes the basic characteristics of the tools analyzed.

Classification of items in the appraisal tools

Following the framework proposed by Furlong and Oancea (2005), the coding outcomes of the 12 selected tools were as follow: A majority (99%) of the appraisal items (192 items) were related to the Epistemic domain which represents their methodological and theoretical robustness; 2 items were related to the Technological domain; no item was related to the Value for people domain nor the Economic domain. Among the 12 appraisal tools, none of them covers all the four domains of Furlong's framework; only EPPI cover the first two domains; the other 11 tools cover the epistemic domain only. Amongst the 192 items in the Epistemic domain, 164 items were categorized into the sub-domain of Trustworthiness; 24 items were related to the sub-domain of Explicitness in Design and Reporting; 3 items were in the sub-domain of Propriety, and 1 item was related to sub-domain of Paradigm-dependent Criteria.

When further look into the sub-domain of trustworthiness, as shown in Table 2, the 164 items in Trustworthiness fall into the six broad methodological categories: (i) Sample and Sampling --

31 items; (ii) Data Collection -- 48 items; (iii) Intervention -- 22 items; (iv) Data Analysis -- 22 items; and (v) Findings and Conclusions -- 12 items; while (vi) 29 items measure "Overall Quality" (Table 2).

Discussion

In this study, the search, selection and examination of appraisal tools used in educational research provide an overview of the status quo of appraisal tools used in the systematic review of educational research. This will serve as a guide to help education researchers to select appropriate tools that suit their need in assessing methodological quality of educational studies.

The first domain (Epistemic-Methodological and Theoretical Robustness) Furlong and Oancea (2005) framework consists of sub-domains Trustworthiness, Contribution to Knowledge, Transparency, Explicitness of Design and Reporting. The second domain (Technological) consists of sub-domains "Fitness to Purpose" which include "Purposiveness", "Specificity" and "Accessibility". All items in the 12 appraisal tools were found to fall into these two domains and associated sub-domains. This may lead to a query whether and why other domains and sub-domains (related to whether the research facilitates wise and reflective practice and examples of cost-effectiveness and marketability) are not considered major concerns when educational research is critically appraised.

To be more specific, results of the coding revealed that most of the items in the tools fall into the sub-domain Trustworthiness. There may be a need for additional works on the development of appraisal items focusing on other dimensions (for example, the Capacity Development and Value for People Domain and the Economic Domain) for a more comprehensive review of which research has passed a specified threshold of quality.

On the other hand, as most of the existing appraisal tools in the educational field have been used in quantitative studies, the review and development of appraisal tools for qualitative studies require further research. Nonetheless, some classical definition [7, 8] of Trustworthiness included the context of validity, reliability, groundedness, dependability and believability. This seems to form an argument that the two major and seemingly different approaches, quantitative and qualitative, are not necessarily mutually exclusive which is worth further debate and examinations on how they complement each other in ways that have not been explored before.

TABLE 1 -- BASIC FEATURES OF THE 12 SELECTED APPRAISAL TOOLS

	Appraisal Tool	Country	Year	Area originally design for	Type of method	No. of items on methodological quality included in this study
1	EPPI	UK	2007	Education	Mixed	16
2	CRD	UK	1994	Health & Medicine	Mixed	7
3	WWC	USA	2002	Education	Quantitative	6
4	BEE	UK	2008	Education	Quantitative	8
5	CSRQ	USA	2002	Education	Quantitative	22
6	JBI-QARI	Australia	2007	Health & Medicine	Qualitative	9
7	Cochrane Collaboration	USA	2008	Health & Medicine	Quantitative	5
8	Quality Assessment Scale	USA	2007	Health & Medicine	Quantitative	10
9	Study DIAD	USA	2007	Social Sciences	Mixed	35
10	Quality Indicators	USA	2005	Social Sciences	Quantitative	14
11	CEBP Checklist	USA	2010	Education	Quantitative	15
12	Checklist for Study Quality	UK	1998	Social Sciences	Quantitative	17

TABLE 2 -- CATEGORIZATION OF THE ITEMS IN THE 12 APPRAISAL TOOLS

	Category	No. of Items Included (Total 164)	Details and Remarks
1	Sample & Sampling	31 Items	Further grouped into 3 categories: (i) Sample equivalence, baseline and pre-test differences (8 items); Randomization (12 items); (iii) Sample size, representativeness (11 items)
2	Data Collection	48 Items	Further grouped into 4 categories: (i) Attrition (5 items); (ii) Blinding (6 items);(iii) Systematic and empirical data collections (13 items); (iv) Outcome measures (24 items)
3	Intervention	22 Items	(i) Design and conduct of intervention (22 items)
4	Data Analysis	22 Items	Further grouped into 2 categories:(i) Rigorous data analysis (17 items); (ii) Adjustment of confounding (5 items)
5	Findings & Conclusion	12 Items	Further grouped into 2 categories: (i) Effect Size (7items); (ii) Internal consistency (5 items).
6	Overall Quality	29 Items	Further grouped into 3 categories: (i) Overall weight of evidence (9 items); (ii) Bias (7 items); (iii) External validity (13 items)

REFERENCES

1. O. Wendt, and B. Miller. Quality appraisal of single-subject experimental designs: An overview and comparison of different appraisal tools. *Education and Treatment of Children.* 2012, 35(2): 235-268.

2. M. Petticrew, and H. Roberts. *Systematic Reviews in the Social Sciences: A Practical Guide.* Oxford: Blackwell, 2006.

3. H. Ader, G. Mellenbergh, D. Hand. Methodological quality (chapter 3) in *Advising on Research Methods: A consultant's companion*. Johannes van Kessel Publishing, 2008.

4. J. Thomas, A. Harden, A. Oakley, S. Oliver, K. Sutcliffe, R. Rees, G. Brunton, J. Kavanagh. Integrating qualitative research with trials in systematic reviews. BMJ. 2004, 328 (7446): 1010–1012.

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

6. J. Furlong, and A. Oancea. A. Assessing quality in applied and practice-based educational research: A framework for discussion. *Review of Australian Research in Education: Counterpoints on the Quality and Impact of Educational Research.* 2005, 6: 89-104.

7. A. Strauss, and J, Corbin. *Basics of Qualitative Research. Grounded Theory Procedures and Techniques.* Newbury Park, Sage, 1990.

8. Y. S. Lincoln, and E. G. Guba. *Naturalistic Inquiry.* Newbury Park, London, New Delhi: SAGE Publications, 1985.