

Please cite this article as:

Heyvaert, M., Hannes, K., Maes, B., & Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, 7(4), 302-327.

doi:10.1177/1558689813479449

Critical Appraisal of Mixed Methods Studies

Abstract

In several subdomains of the social, behavioral, health, and human sciences, research questions are increasingly answered through mixed methods studies, combining qualitative and quantitative evidence and research elements. Accordingly, the importance of including those primary mixed methods research articles in systematic reviews grows. It is generally known that the critical appraisal of articles is an essential step in the development of a methodologically sound review. This paper provides an overview of the available critical appraisal frameworks developed to evaluate primary mixed methods research articles. In addition, we critically compare and evaluate these frameworks and the quality criteria they include.

Key words: critical appraisal; quality assessment; quality evaluation; mixed methods research; systematic review; research synthesis

Introduction

In the last decades the popularity of mixed methods research (MMR) has increased steadily (Creswell, 2003; Greene, 2007; Johnson & Onwuegbuzie, 2004; Onwuegbuzie & Leech, 2005a; Tashakkori & Creswell, 2007; Tashakkori & Teddlie, 2003a). Studies combining qualitative and quantitative research elements are now regularly conducted in several subdomains of the social, behavioral, health, and human sciences (Alise & Teddlie, 2010; Collins, Onwuegbuzie, & Sutton, 2006; Creswell, 2003; Fidel, 2008; Greene, 2007; Hart, Smith, Swars, & Smith, 2009; Hurmerinta-Peltomaki & Nummela, 2006; Hutchinson & Lovell, 2004; Johnson & Onwuegbuzie, 2004; Tashakkori & Creswell, 2007; Truscott et al., 2010). Accordingly, this type of primary evidence article represents a substantial part of the available evidence concerning multiple research topics in those research domains. As a consequence, authors involved in systematically reviewing scientific literature on one of these topics are challenged to include this type of study in their reviews.

Several authors (e.g., Cooper, 2010; Cooper, Hedges, & Valentine, 2009; Higgins & Green, 2009; Petticrew & Roberts, 2006) as well as organizations such as the Cochrane Collaboration and the Campbell Collaboration have produced clear guidelines on how to conduct each step of a systematic review of quantitative and qualitative evidence, more specifically on searching

strategies, critical appraisal strategies, data extraction, and the synthesis of the findings of the included primary studies. However, relatively little effort has been made to stipulate how exactly to deal with mixed methods (MM) primary research in the predefined methodology of systematic reviews.

Recently, some pioneering work has been conducted on synthesizing findings from a variety of qualitative, quantitative, and MM primary research evidence in systematic reviews applying MMR approaches (see Heyvaert, Maes, & Onghena, 2011a, for a comprehensive overview). Developed frameworks for undertaking such reviews include: *integrative review* (Whittemore & Knafl, 2005), *meta-needs assessment* (Gaber, 2000), *mixed methods research synthesis* (Heyvaert, Maes, & Onghena, 2011b), *mixed methods synthesis* (Harden & Thomas, 2005, 2010), *mixed research synthesis* (Sandelowski, Voils, & Barroso, 2006), *mixed studies review* (Pluye, Gagnon, Griffiths, & Johnson-Lafleur, 2009b), and *realist review* (Pawson, Greenhalgh, Harvey, & Walshe, 2005). Answers to the questions how a systematic review applying MMR approaches can be conducted and which designs are suitable for these reviews can be found in the work of Sandelowski et al. (2006), Heyvaert et al. (2011b), and Onwuegbuzie, Collins, Leech, Dellinger, and Jiao (2010).

The critical appraisal of articles to be included in a review is an essential step in the development of a methodologically sound review (Cooper, 2010; Khan, Riet, Popay, Nixon, & Kleijnen, 2001; Major & Savin-Baden, 2010; Whittemore & Knafl, 2005). After all, it is generally known that the inclusion of studies that are methodologically flawed could lead to a substantial bias in the end result of a systematic review (Cooper & Hedges, 1994; Higgins et al., 2008). That is why several authors and institutes have developed frameworks to evaluate the methodological quality of primary qualitative and quantitative articles (<http://www.unisa.edu.au/cahe/Resources/CAT/default.asp>). However, critical appraisal frameworks (CAFs) developed to evaluate the methodological quality of primary MMR articles are less commonly found in scientific literature.

The authors that developed frameworks for undertaking systematic reviews applying MM approaches (e.g., Gaber, 2000; Harden & Thomas, 2005, 2010; Heyvaert et al., 2011b; Onwuegbuzie et al., 2010; Pawson et al., 2005; Pluye et al., 2009b; Sandelowski et al., 2006; Whittemore & Knafl, 2005) do mention the necessity of critically appraising the methodological quality of all the included primary research articles; however, they only seldom indicate *which* quality criteria can be applied to primary MMR articles. Because a

MM study is more than just the sum of the individual qualitative and quantitative strands of the study (Day, Sammons, & Gu, 2008; Hall & Howard, 2008; Moffatt, White, Mackintosh, & Howel, 2006; Tashakkori & Teddlie, 2003b), the combined application of qualitative and quantitative critical appraisal criteria is most likely not sufficient to evaluate the methodological quality of a primary MMR article.

This paper provides an overview of the available CAFs developed to evaluate the methodological quality of primary MMR articles. Additionally, we want to compare critically and evaluate the quality criteria proposed in these frameworks.

The authors of the present study take pragmatism as their MM philosophical stance, advocating the integration of quantitative and qualitative methods within a single study from a question-driven philosophy (cf. Heyvaert et al., 2011b; Onwuegbuzie & Leech, 2005b). This implies that one should apply the best suited combination of methods and modes of analysis in order to answer the posed research question(s): that can be a monomethod or a MM approach. Emphasizing processes of abduction, intersubjectivity, and transferability (Morgan, 2007), pragmatism offers the researcher alternatives to the dichotomous choice between (post)positivism and constructivism, driven by the question of utility (cf. Biesta, 2010; Creswell & Plano Clark, 2007; Feilzer, 2010; Hannes & Lockwood, 2011; Morgan, 2007).

Search strategy for identification of studies and inclusion criteria

We searched for manuscripts published up to December 31, 2009, reporting on CAFs developed to evaluate the methodological quality of primary MMR articles. December 31, 2009 was used as the cut-off point for the search since the data collection started in January, 2010. The following databases were searched: *Academic Search Premier (ASP)*, *Allied and Complementary Medicine (AMED)*, *British Education Index*, *Cumulative Index to Nursing and Allied Health Literature (CINAHL)*, *Embase*, *Education Resources Information Center (ERIC)*, *Francis*, *Medline*, *PsycINFO*, *PubMed*, and *Sociological Abstracts*. The search was extended by focusing on grey literature databases and dissertations and theses databases. In order to trace grey literature and dissertations and theses, 10 databases were searched: the *CORDIS Library*, *Educational Technology and E-Learning (EdITLib)*, the *Grey Literature Database of the Canadian Evaluation Society*, the *Index of Conference Proceedings*, the *Index to Theses in Great Britain and Ireland*, the *International Bibliography of the Social Sciences (IBSS)*, *ProQuest Dissertations & Theses*, the *Social Science Research Network*

(SSRN) eLibrary, the *System for Information on Grey Literature in Europe (SIGLE)*, and *Theses Canada*. In addition, we conducted a hand search of 10 journals with a tradition of providing information on methodology of MMR: *Educational Researcher*, *Field Methods*, *International Journal of Multiple Research Approaches*, *International Journal of Research and Method in Education*, *International Journal of Social Research Methodology*, *Journal of Mixed Method Research*, *Qualitative Inquiry*, *Qualitative Report*, *Quality & Quantity*, and *Research in the Schools*. For this hand search we screened the titles and abstracts from the publications included in these journals. We further searched the reference lists of all identified relevant papers (*backward search*) and retrieved more recent references through searching three citation databases (*forward search*). The three indexes we consulted in order to conduct cited reference searching were the *Arts & Humanities Citation Index*, the *Science Citation Index Expanded (SCI)*, and the *Social Sciences Citation Index (SSCI)*. All three citation indexes are included in the *Web of Science* database. When and where possible we used electronic information on abstracts, instead of reading through all the retrieved articles. Finally, authors of retrieved relevant studies and MMR experts were contacted regarding any additional published or unpublished work.

We conducted the search process by combining search terms related to MM review designs with keywords used to identify CAFs. The former search string included *integrative review*, *meta-needs assessment*, *mixed method(s)*, *mixed methods synthesis*, *mixed research synthesis*, *mixed studies review*, *multi(-)method*, and *realist review*. For the databases *British Education Index*, *CINAHL*, *CORDIS Library*, *Embase*, *ERIC*, *Francis*, *IBSS*, *Index of Conference Proceedings*, *Index to Theses in Great Britain and Ireland*, *International Bibliography of the Social Sciences*, *PubMed*, *Social Science Research Network*, *Sociological Abstracts*, and *Theses Canada* this string was expanded with the search terms *meta-analysis*, *review*, and *synthesis*, because the original design-related search terms retrieved very few manuscripts. Applied search terms related to CAFs were *appraisal framework*, *checklist*, *critical appraisal*, *quality appraisal*, *quality assessment*, and *quality evaluation*. Both search strings were combined by using a Boolean logic.

Only publications reporting on CAFs developed to evaluate the methodological quality of primary MMR articles were included. We did not use any restriction on the language of papers. When separate manuscripts presented a particular version of one CAF, we included the article presenting the most comprehensive framework. In case when separate manuscripts

presented exactly the same framework, we only included the original manuscript.

Screening, extraction, and analysis

The search for manuscripts was conducted by the first author. As an agreement check, all titles and abstracts of the retrieved studies from one randomly chosen (Haahr, 1998) database (*ASP*), grey literature database (*IBSS*), and journal (*Qualitative Inquiry*) were screened for inclusion by a second, independent researcher. In total, 1422 articles were involved in the check for inter-coder agreement. Full text copies of all potentially relevant papers were retrieved. Four remaining disagreements on the relevance for inclusion were resolved by involving a third researcher.

We provide an overview of the available CAFs developed to evaluate the methodological quality of primary MMR articles. We cross-compare and tabulate the criteria that are addressed in the retrieved frameworks. Applying a constant comparative method, we generate headings that group similar criteria of the retrieved CAFs. The categorization involves interpretive and iterative processes. In addition, we critically discuss the retrieved CAFs for primary MMR articles and their included criteria.

Results and discussion

Study retrieval and interrater agreement

Our search for publications reporting on CAFs that evaluate the methodological quality of primary MMR articles, resulted in 3 included publications through the databases search, 1 through the grey literature databases and dissertations and theses databases search, 5 through the hand search of journals, 12 through the screening of the reference lists, and 6 through the search of the citation databases. Finally, we contacted 23 authors of retrieved relevant studies and MMR experts regarding any additional published or unpublished work. From 16 of them we received an answer: 5 authors and experts stated that they did not know of other tools for evaluating the quality of MMR and that they were unaware of any additional publications on this topic, and 11 authors and experts emailed us at least one reference of a publication on this topic that was not yet included in our list of included articles. However, from the 14 publications retrieved this way, 8 were published after December 31, 2009. From the 6 remaining publications, 1 study fitted our inclusion criteria.

Several publications were retrieved by more than one of the above-named search strategies.

Our database included 18 unique publications. However, we additionally excluded five manuscripts from this review, because they described the same CAF as presented in one of the already included studies. The references to those five manuscripts are preceded by ** in the annotated bibliography. As a result, our final database included 13 unique CAFs. The manuscripts containing these frameworks are preceded by * in the annotated bibliography. Results from the search are presented in a flowchart (Figure 1).

As a check for the screening exercise interrater agreement was calculated by dividing the number of agreements on inclusion by the number of agreements plus disagreements. The interrater agreement was 99.93%.

INSERT FIGURE 1 ABOUT HERE

Comparing the CAFs

Figure 2 provides an overview of the 13 included CAFs for evaluating the methodological quality of primary MMR articles (Alborz & McNally, 2004; Bryman, Becker, & Sempik, 2008; Caracelli & Riggin, 1994; Creswell & Plano Clark, 2007; Dellinger & Leech, 2007; Dybå, Dingsøy, & Hanssen, 2007; Greene, 2007; O’Cathain, Murphy, & Nicholl, 2008; Onwuegbuzie & Johnson, 2006; Pluye, Gagnon, Griffiths, & Johnson-Lafleur, 2009a; Pluye, Grad, Dunikowski, & Stephenson, 2005; Sale & Brazil, 2004; Teddlie & Tashakkori, 2009). Using a constant comparative method, we generated 13 headings that group similar criteria of the retrieved frameworks: *criteria for qualitative part of the study; criteria for quantitative part of the study; criteria for mixing and integration of methods; rationale for mixing methods stated; theoretical framework; research aims and questions; design; sampling and data collection; data analysis; interpretation, conclusions, inferences, and implications; context; impact of investigator; and transparency*. The first and second column heading refer to criteria in the frameworks that explicitly suggest to score separately the methodological quality of the qualitative and quantitative strands of a study. The third and fourth column contain criteria that are explicitly concerned with MMR: the mixing and integration of the combined methods and strands, and rationales for conducting MMR. The nine other column headings concern generic criteria, that are also often included in tools for critically appraising qualitative primary research studies and in tools for critically appraising quantitative primary research studies (e.g., Letts et al., 2007; Public Health Resource Unit, 2006a, 2006b): (a) stating the theoretical framework of the study; (b) stating the research aims and questions; (c)

using an appropriate design; (d) applying appropriate sampling and data collection methods; (e) applying appropriate data analysis methods; (f) stating the interpretation, conclusions, inferences, and implications of the study at hand; (g) stating the context of the research; (h) stating the impact of the researchers; and (i) being transparent in the reporting of the study. The nine headings refer to an important primary study element that should be clearly described in a research report, and for which one must judge whether its realization is adequate in the present study. Figure 2 contains the criteria that are addressed in the retrieved frameworks. For the detailed descriptions of these criteria, and for indicators of the criteria, we refer the reader to the original frameworks (Alborz & McNally, 2004; Bryman et al., 2008; Caracelli & Riggin, 1994; Creswell & Plano Clark, 2007; Dellinger & Leech, 2007; Dybå et al., 2007; Greene, 2007; O’Cathain et al., 2008; Onwuegbuzie & Johnson, 2006; Pluye et al., 2005, 2009a; Sale & Brazil, 2004; Teddlie & Tashakkori, 2009). Table 1 provides an overview of the criteria included in the 13 retrieved CAFs, with the frequency that each criterion is mentioned in the frameworks.

INSERT FIGURE 2 ABOUT HERE

INSERT TABLE 1 ABOUT HERE

Retrieving 13 unique CAFs developed to evaluate the methodological quality of primary MMR articles indicates that several authors have been working on this domain. Because 12 out of 13 frameworks have been published between 2004 and 2009, the question of how to appraise the methodological quality of primary MMR articles has emerged as being a very contemporary one. However, standard protocols turn out to be lacking, and substantial differences between the construction of the 13 frameworks and the included quality criteria can be observed.

Specific critical appraisal criteria for the qualitative and quantitative strands of a MMR study

Regarding the first and second column of Figure 2, we notice that nine frameworks suggest to score separately the methodological quality of the qualitative and quantitative strands of a study. Accordingly, Bryman (2006a) calls this the *separate criteria* approach. Most authors formulate criteria for specifically judging the qualitative and quantitative strands of a study in a generic way (Alborz & McNally, 2004; Dellinger & Leech, 2007; Greene, 2007; O’Cathain

et al., 2008; Sale & Brazil, 2004; Teddlie & Tashakkori, 2009). Pluye et al. (2005, 2009a) propose design-dependent criteria.

Specific critical appraisal criteria for MMR

We identified two groups of criteria that are explicitly concerned with MMR: the mixing and integration of the combined methods and strands, and providing a rationale for conducting MMR. Bryman (2006a) uses the term *bespoke criteria* for quality appraisal criteria that are especially devised for MMR. Concerning the third column of Figure 2, nine frameworks explicitly include criteria pertaining to the mixing and integration of methods. Several of the retrieved frameworks simply include the question whether the qualitative and quantitative strands of the studies are actually integrated, and whether this integration is carried out adequately (Alborz & McNally, 2004; Bryman et al., 2008; Creswell & Plano Clark, 2007; Pluye et al., 2009a). Sale and Brazil (2004) stress four general criteria to appraise critically primary MM studies that refer to Lincoln and Guba's (1985, 1986) framework of trustworthiness and rigor: truth value, applicability, consistency, and neutrality. O'Cathain et al. (2008) pose nine questions for assessing the integration of MM studies, on the type of integration and its appropriateness to the design, rigor, and the time allocation and team work concerning the integration. Teddlie and Tashakkori (2009) introduce the concept of integrative efficacy as criterion pertaining to the mixing and integration of methods, questioning whether meta-inferences adequately incorporate the inferences made in each strand of the study, and whether theoretical explanations are offered when inconsistencies exist between the inferences made (this criterion is presented as part of *interpretive rigor* in their framework). Onwuegbuzie and Johnson (2006) describe nine types of legitimation (cf. Figure 2) as standards for assessing the quality of the mixing and integration of methods. Dellinger and Leech (2007) refer in their validation framework to the criteria proposed by Onwuegbuzie and Johnson (2006) and Teddlie and Tashakkori (2003, 2009) and propose no addenda or adjustments hereto, which indicates that they thoroughly agree with the criteria proposed by these authors.

Looking at the fourth column of Figure 2, we see that the criterion to report the rationale for the applied MMR approach is included in four of the retrieved frameworks. The criterion implies that researchers provide a clear and defensible justification for mixing qualitative and quantitative approaches. Making explicit the rationale for conducting a MMR study stimulates a thoughtful decision process concerning the design and implementation of the study (cf. Collins et al., 2006). An influential framework of rationales for MMR is that of Greene,

Caracelli, and Graham (1989). Based on a theoretical review, they identified five broad rationales of MMR studies: triangulation, complementarity, development, initiation, and expansion (pp. 258-260). More recently, Collins et al. (2006) listed rationales for MMR proposed by various authors (pp. 74-75), and grouped them into four major rationales: participant enrichment, instrument fidelity, treatment integrity, and significance enhancement (pp. 76-89). Also in 2006, Bryman (2006b) conducted a content analysis of 232 MMR studies and studied the rationales that are given for using a MMR approach. Therefore, he devised a scheme consisting of 16 rationales: triangulation, offset, completeness, process, different research questions, explanation, unexpected results, instrument development, sampling, credibility, context, illustration, utility, confirm and discover, diversity of views, and enhancement (pp. 105-107).

Generic critical appraisal criteria

The nine other column headings of Figure 2 concern generic critical appraisal criteria. In critical appraisal tools for qualitative as well as for quantitative primary studies (see Introduction), these nine criteria are often applied to appraise the epistemological and methodological rigor of a study in a more generic way. Although it is a prerequisite of any research report to state clearly all the applied data analysis procedures, only 7 of the 13 retrieved frameworks explicitly contain this criterion. The criterion to delineate the research aims and questions is only included in five frameworks. The criterion to report the applied sampling and data collection procedures is only included in four of the retrieved frameworks. Likewise, the criteria that in a primary MMR study the theoretical framework, and the impact of investigator on the research process and product should be clearly reported, is only mentioned in 3 of the 13 retrieved frameworks. Moreover, the criterion to include a plain description of the context of the study and the criterion to be transparent about the employed procedures are only included in two frameworks. An explanation for the relative scarce prevalence of these criteria is that although some authors mention these criteria when describing indicators or giving a description of a broad range of design-related issues in general, they often do not state them as separate criteria. For example, concerning the criterion *is the applied design appropriate?*, several diverging indicators for this criterion can be included, pertaining to the data analysis (e.g., *does the design include appropriate data analysis procedures?*), the research aims and questions (e.g., *does the design match the stated research aims?*), the theoretical framework, the applied sampling and data collection procedures, and so on. It is possible that the authors of the retrieved frameworks did not

separately mention these generally accepted criteria, because they do not especially apply to MM studies. However, incorporating too extensive descriptions of criteria in a CAF brings along the risk that reviewers appraising studies by means of such a framework find it very difficult to judge whether all indicators of a certain criterion are sufficiently elaborated and reported in order to judge whether this *entire* criterion is adequately addressed. A framework developed for the evaluation of the methodological quality of primary MMR articles should contain criteria that have a limited number of clear-cut indicators, in order to facilitate the critical appraisal work of its user.

Two of the nine generic critical appraisal criteria are respectively included in nine and eight retrieved frameworks. The first concerns the design, the second concerns the interpretations, conclusions, inferences, and implications of the study at hand. With regard to the *design* of the study, several frameworks simply include the question whether the design is clearly described, and whether it is appropriate to the research aims (Alborz & McNally, 2004; Creswell & Plano Clark, 2007; Dybå et al., 2007; Pluye et al., 2009a). Caracelli and Riggin (1994) add specific questions on the triangulation design, on the combination of strengths and weaknesses of different methods, and on the minimalization of shared bias between methods. O’Cathain et al. (2008) describe additional questions on the feasibility and success of the design, and on its rigor. Teddlie and Tashakkori (2009) group design quality indicators together under three criteria: design suitability (appropriateness), design fidelity (adequacy), and within-design consistency (their criterion *analytic adequacy* is put under the heading *data analysis* in Figure 2). By doing that, they propose a useful classification for design-related quality criteria in which the listed design-related quality criteria from the five other frameworks could be positioned. Their classification has been used by Dellinger and Leech (2007) without further adaptation.

With regard to the column heading *interpretations, conclusions, inferences, and implications*, we notice that although some authors describe a wide range of criteria related to this category, 5 out of the 13 retrieved frameworks did not separately contain criteria related to this category. Although it was an option to differentiate further within this category, we chose not to, because criteria relating to interpretations, conclusions, inferences, and implications of the research were often presented jointly in the retrieved frameworks, and parting them would lead to a distorted picture that does not correspond to the original frameworks on appraising MM studies. Creswell and Plano Clark (2007) pose in their framework the general question

whether good conclusions or inferences are drawn, while Dybå et al. (2007) additionally pose questions pertaining to the clarity of statement of the findings (with credible results and justified conclusions) and to the statement of the study's value for research and practice. O'Cathain et al. (2008) add considerations on clarity about which results have emerged from which methods, on the appropriateness of the inferences, and on considering the results of all the applied methods in the interpretation. On a more abstract level, Greene (2007, p. 167) argues that for warranting the quality of inferences, conclusions, and interpretations made, a multiplistic stance should be adopted, that (a) focuses on the available data support for the inferences, using data of multiple and diverse kinds; (b) could include criteria or stances from different methodological traditions; (c) considers warrants for inquiry inferences a matter of persuasive argument, next to a matter of fulfilling established criteria; and (d) attends to the nature and extent of the better understanding that is reached with this MM design. Again introducing a distinct terminology, with respect to criteria on interpretations, conclusions, inferences, and implications of the research, Teddlie and Tashakkori (2009) refer to *interpretive rigor* and *inference transferability*. The term *interpretive rigor* encompasses the criteria interpretive consistency (consistency of inferences with relevant findings, consistency between inferences), theoretical consistency (consistency of inferences with theory), interpretive agreement (agreement of other scholars and participants with the conclusions), interpretive distinctiveness (credibility and plausibility of the inferences made), and interpretive correspondence (correspondence of the inferences to the purposes of the study and of the design). Dellinger and Leech (2007) again include the terminology proposed by Teddlie and Tashakkori into their validation framework, and now add considerations on translation fidelity and inferential consistency, together with questions on the utilization and consequences of the findings or measures. Finally, Caracelli and Riggin (1994) list criteria on a study's interpretations, conclusions, inferences, and implications that especially concern issues such as assigned weights, biases, comprehensiveness, interpretability, and value of the study for stakeholders and policy.

In need of distinct CAFs for MMR studies

Of the 13 headings grouping similar criteria of the retrieved frameworks, 2 refer to criteria for scoring separately the methodological quality of the qualitative and quantitative strands of a study, 2 refer to criteria that explicitly concern MMR (i.e., the mixing and integration of the combined methods and strands, and providing a rationale for conducting MMR), and 9 others

refer to generic critical appraisal criteria that are also often included in tools for critically appraising qualitative and quantitative primary research studies.

Merely assessing the individual qualitative and quantitative strands (columns 1 and 2 in Figure 2) when critically appraising MM studies is too limited, because a MM study is more than simply the sum of its qualitative and quantitative elements (see Creswell & Plano Clark, 2011; O’Cathain, 2010). We present a few reasons why, along with appraising the specific strands of a MM study, the methodological quality of the MM analysis and inferences overarching the qualitative and quantitative strands should be considered when including a MM study in a systematic review. First, the overarching MM analysis can inform the reader whether and concerning what aspects the qualitative and quantitative evidence accords or conflicts. By mixing these strands in an overarching MM analysis, strengths of one approach can compensate for weaknesses of the other approach, resulting in a firmer confirmation of a theory when the diverse evidence accords. When the evidence conflicts, the overarching MM analysis can critically interrogate the applied methods, or highlight different aspects of the research question that are separately answered by the qualitative and quantitative evidence. Second, when one strand informs decisions for designing a second strand (e.g., when a preceding qualitative strand sparks new hypotheses concerning a group of outlying cases that can be tested in a second, quantitative strand), the strands cannot be appraised as independent, and the overarching MM design and analysis should be taken into account when appraising the study, next to appraising its specific strands. Third, when complementary insights from both strands together create a bigger picture and answer different subquestions of one overarching research question (e.g., *what is it about this kind of intervention that works, for whom, in what circumstances, in what respects, and why?*; Pawson et al., 2005), the quality of the overarching MM analysis matters to a systematic reviewer, next to the quality of its qualitative and quantitative strands.

Following the definitions of MMR offered by leading scholars (discussed in Johnson, Onwuegbuzie, & Turner, 2007), the aspect most characteristic to MMR is the mixing or combining of qualitative and quantitative research elements. Additionally, these authors stress the importance of rationales of MMR studies. Accordingly, the criteria *appropriateness of mixing qualitative and quantitative research components* and *providing a rationale for conducting MMR* should be included in each CAF for evaluating MMR articles (columns 3 and 4 in Figure 2).

Concluding, we state that authors who systematically review scientific literature including MMR studies should appraise those studies with a critical appraisal instrument specifically

designed for MMR studies: merely applying critical appraisal instruments for the separate qualitative and quantitative strands can not suffice (cf. Creswell & Plano Clark, 2011; O’Cathain, 2010), nor can the use of a *generic* critical appraisal instrument suffice (cf. Katrak, Bialocerkowski, Massy-Westropp, Kumar, & Grimmer, 2004; Young & Solomon, 2009), that does not stress the importance of appropriately mixing the qualitative and quantitative research elements and of providing a rationale for conducting MMR. The qualitative and quantitative strands of a MM study should not only be answering to strand-specific criteria; in addition, the strands should be appropriately mixed in order to answer the posed research questions, a rationale for the MMR approach should be provided, and the overall study should be coherent and insightful.

Evaluation of the criteria included in the CAFs

Our analysis of the retrieved CAFs was an inductive one, generating headings that group similar criteria of the retrieved frameworks by applying a constant comparative method. Additional to this inductive approach, it could be interesting to compare the CAFs to a prior frame enumerating the most important issues about MMR, in order to see whether the CAFs cover these issues well, or if there are gaps. The frame of Creswell (2010) is applied as reference frame. It encompasses five domains: (a) the essence of MM domain, (b) the philosophical domain, (c) the procedures domain, (d) the adaptation and use domain, and (e) the political domain. This frame is based on the work of Creswell (2008, 2009), Greene (2008), and Tashakkori and Teddlie (2003b).

(a) *The essence of MM domain* concerns issues like MMR nomenclature and the nature of MMR. The only retrieved CAF explicitly including criteria belonging to this domain is that of Creswell & Plano Clark (2007) (*using MM terms to describe the study*).

(b) *The philosophical domain* encompasses paradigmatical MMR issues. CAFs explicitly including criteria belonging to this domain, are those of Creswell and Plano Clark (2007) (*acknowledging the paradigm stance of the researcher*); Dellinger and Leech (2007) (*paradigmatic mixing*); Greene (2007) (*mixing paradigms and mental models*); and Onwuegbuzie and Johnson (2006) (*paradigmatic mixing*). This domain is further discussed in the next section.

(c) *The procedures domain* includes research design issues, validity and evaluation issues, inquiry logic issues, techniques of MMR, and providing a rationale for MMR. *All* retrieved CAFs include criteria belonging to this domain.

(d) *The adaptation and use domain* concerns topics like collaborating on MMR projects, teaching MMR, reporting MMR, disciplinary developments, and international growth. CAFs explicitly including criteria belonging to this domain, are those of Bryman et al. (2008) (*transparency*); Caracelli and Riggin (1994) (*cluster 7: reporting criteria*); Creswell and Plano Clark (2007) (*reporting detailed quantitative and qualitative procedures*); Dellinger and Leech (2007) (*utilization/historical element*); Greene (2007) (*considering warrants for inquiry inferences a matter of persuasive argument as well as fulfilling criteria*); O’Cathain et al. (2008) (*criteria on the success of the MM study, team-related criteria on the integration in the MM study*); and Onwuegbuzie and Johnson (2006) (*team-related inside-outside legitimation issues*).

(e) *The political domain* encompasses questions about the audience, the represented perspective, the voices heard, who is being advocated for, funding opportunities, and about justifying MMR. CAFs explicitly including criteria belonging to this domain, are those of Alborz and McNally (2004) (*policy relevance*); Caracelli and Riggin (1994) (*cluster 6: stakeholders criteria*); Dellinger and Leech (2007) (*consequential element*); Dybå et al. (2007) (*studies’ value for research and practice*); Greene (2007) (*social inquiry guided by practical philosophy*); and Onwuegbuzie and Johnson (2006) (*political legitimation*).

Summarizing, we see that all five domains listed by Creswell (2010) are covered by at least one criterion of one of the retrieved CAFs. However, there are large differences between the number of domains included in the respective frameworks. None of the retrieved framework yet encompasses all domains stipulated by Creswell (2010). There are four frameworks that include criteria relating to four of the five domains: the CAFs of Creswell and Plano Clark (2007), Dellinger and Leech (2007), Greene (2007), and of Onwuegbuzie and Johnson (2006). So, judging based on the frame of Creswell (2010), these 4 CAFs are the most versatile of the 13 retrieved frameworks.

There are large differences between the domains concerning how often they are referred to in the retrieved CAFs. For all CAFs, the *procedures domain* was the most extensively described domain within the lists of criteria. The *political domain* and the *adaptation and use domain* were included in criteria of respectively six and seven CAFs. Criteria relating to the *philosophical domain* were explicitly included in four CAFs (see also next section). There was only one framework that included criteria on the *essence of MM domain*.

Philosophical stances

When interpreting the developed CAFs for evaluating the methodological quality of primary MMR articles, an important factor to consider is the philosophical stance underlying each framework. Depending on what a CAF developer's philosophical stance is (implicitly or explicitly), he will 'value' different things in a MMR study, and develop a CAF accordingly. Greene and Hall (2010) enumerate five stances on mixing paradigms while mixing methods: (a) purist, (b) complementary strengths, (c) dialectic, (d) aparadigmatic, and (e) pragmatism as alternative paradigm. Researchers within these stances hold different answers to the question *what is the importance and role of philosophical assumptions in inquiry practice?*: researchers within the first three stances say this is highly important, aparadigmatic researchers state this is not really important, and pragmatic researchers give answers ranging from *highly* to *not really important* (Greene & Hall, 2010, p. 123). We expect that authors believing that philosophical assumptions are highly important in inquiry practice, would at least include a criterion such as *Is the paradigm stance of the researcher acknowledged?* in their CAF.

Four out of the 13 CAFs do not include criteria related to philosophical assumptions (Bryman et al., 2008; Dybå et al., 2007; Pluye et al., 2005, 2009a), and tend to be aparadigmatic frameworks. Five CAFs do not explicitly include criteria related to philosophical assumptions, but do mention the importance of philosophical assumptions in the text accompanying the CAF, e.g. by making the suggestion to note down comments on the used paradigm and MM approach for each retrieved primary article (Alborz & McNally, 2004; Caracelli & Riggin, 1994; O'Cathain et al., 2008; Sale & Brazil, 2004; Teddlie & Tashakkori, 2009). These frameworks tend to be pragmatic frameworks with a strong instrumental orientation and a smaller orientation to philosophical assumptions (cf. Greene & Hall, 2010). Four CAFs do explicitly include criteria related to philosophical assumptions (Creswell & Plano Clark, 2007; Dellinger & Leech, 2007; Greene, 2007; Onwuegbuzie & Johnson, 2006). The frameworks of Creswell and Plano Clark (2007), Dellinger and Leech (2007), and Onwuegbuzie and Johnson (2006) are pragmatic frameworks with a stronger orientation to philosophical assumptions than the former five CAFs (cf. Greene & Hall, 2010; Johnson & Onwuegbuzie, 2004). Taking a different stance, Greene is a strong advocate of the dialectical view, and this reflects in her CAF that lists philosophical assumptions, context, and theory to direct inquiry decisions (cf. Greene, Benjamin, & Goodyear, 2001; Greene & Caracelli, 2003; Greene & Hall, 2010).

A second important factor to consider is the philosophical stance of the users of CAFs, such as researchers engaged in secondary research (i.e., reviews). From our pragmatic perspective (cf. supra) we argue that the choice for using certain CAFs should be based on the ‘utility’ and ‘fit for purpose’ of the CAFs for the studies to be included in the reviews: reviewers should select CAFs that are suitable for the retrieved primary MMR studies (i.e., ‘fit’ between the CAFs and the primary studies to be appraised). For instance, a MMRS-researcher who is conducting a MMRS of primary transformative-emancipatory articles (cf. Mertens, 2010, 2012; Mertens, Bledsoe, Sullivan, & Wilson, 2010) would use a CAF including transformative-emancipatory criteria such as: Do the authors openly reference a problem in a community of concern? Do the authors openly declare a theoretical lens? Were the research questions or purposes written with an advocacy stance? Did the literature review include discussions of diversity and oppression? Did the authors discuss appropriate labeling of the participants? Did data collection and outcomes benefit the community? Were the stakeholders involved in the research project? Did the results elucidate power relationships? Did the results facilitate social change, and were the stakeholders empowered as a result of the research process? Did the authors explicitly state their use of a transformative framework? (cf. Saini & Shlonsky, 2012, pp. 133-135; Sweetman, Badiee, & Creswell, 2010, pp. 442-443). In addition to this pragmatic ‘fit for purpose’ argument, the choice for the CAFs can, implicitly or explicitly, be influenced by the philosophical stance of the researchers doing the reviews. Just as CAF developers may adhere to a philosophical stance, so too may secondary researchers. Based on their philosophical stance, they can ‘value’ some CAFs, and some quality criteria included in CAFs, more than others. For instance, a critical realist researcher who highly values the validity of primary studies would prefer a CAF including criteria such as ‘all statements should be well-grounded in the data’ and ‘the impact of the investigator on the study results should be reported’. However, an interpretivist researcher would subscribe to the argument that the impact of the researcher on the research is inherent to the way research is conducted. The interpretivist researcher may prefer a CAF evaluating issues such as ‘thick description’ and ‘innovative nature of the findings’.

Let us consider three options concerning the use of CAFs. First, one could argue for using a universal CAF, that should be applied to *all* primary MMR studies. An advantage of using a single tool is the comparability of the critical appraisal scores or judgments across all studies. However, considering the divergence of published primary MMR studies, it might not be desirable to apply the same critical appraisal criteria to appraise all primary MMR studies. For instance, in justice oriented participatory MMR research it is a key feature to involve

stakeholders in the research project, but this feature will not be included in a universal MMR CAF. Additionally, considering the divergence of existing philosophical MMR stances, it is not realistic that each secondary researcher would value and apply the same critical appraisal criteria. A second option is to provide a set of general criteria that should minimally be addressed in each primary MMR study, but to additionally allow a secondary researcher to add criteria to this set based on the type of the primary MMR studies to be appraised and/or based on his philosophical stance. So, this second option combines a set of ‘universal’ criteria that should be applied to all MMR studies with specific study-dependent and/or philosophical-stance-dependent criteria. A third option is a ‘pick-and-choose’ CAF: the secondary researcher himself can compose a set of criteria (from a larger pool of criteria) and account for this set based on the type of the primary MMR studies to be appraised and/or on his philosophical stance. As such, this third option does not imply a set of general criteria that should be applied to all MMR studies, but only a set of specific criteria selected by the reviewer. With this third option, it is for instance possible that a reviewer only picks substantive criteria, and leaves methodological appraisal criteria out of his tool.

Construction of CAFs

Concerning the construction of a CAF for the evaluation of the methodological quality of primary articles, we first note that the criteria included in the framework should be mutual exclusive and ought to reflect the most important aspects of the quality of the studies that they assess. Second, a user-friendly critical appraisal instrument should be easy, quick, and clear in its use. Therefore, it should be modeled as an instrument containing a limited number of criteria, with a restricted number of clearly described indicators for each criterion. The list of criteria and indicators should not be too extensive, because that makes the act of appraising the primary studies too time consuming for the reviewer. Additionally, the guidelines for judging or grading the methodological quality of a MM study should be as unambiguous as possible. Some might opt for a critical appraisal tool that assigns numerical ratings according to the methodological quality of a study. Others might prefer a framework that guides a narrative judgment of a primary study at hand. It should be clearly stated how to judge whether separate indicators of a certain criterion are sufficiently addressed in order to evaluate whether an *entire* criterion is adequately addressed. Additionally, it should be clear how to move from a criteria-based approach to a global evaluation of the methodological quality of a study at hand. Instead of aiming at an all-round framework (cf. first option described in *Philosophical stances*), an appealing option could be to provide for each criterion a set of

indicators that should minimally be addressed in the primary study, and to provide accordingly for the study itself a set of criteria that should minimally be addressed (cf. second option described in *Philosophical stances*). It should be clearly stated as well what to do with the final assessment of the study. Should studies that do not reach an optimal score be excluded from the review? Should the effect of dropping studies that do not reach this optimal score be studied in a sensitivity analysis? Furthermore, a critical appraisal tool should document evidence regarding its psychometric properties (score validity and score reliability).

User-friendliness of the retrieved CAFs

When we examine the retrieved frameworks, we have to comment that there is substantial variability in the content and the construction of the frameworks. MMR is a topic that just gained a lot of attention during the last two decades, and only during the last decade the question of how to appraise the methodological quality of primary MMR articles has been addressed by several researchers. Apparently this field is rather novel, and answers to this question of how to appraise MMR articles have been formulated in different forms: as discussion papers, as validation frameworks, as questionnaires for quality assessment without guidelines for judging or grading the studies' quality, and as operational appraisal checklists incorporating rules for grading the studies' quality. There is not yet a consensus on the criteria that should be used to evaluate the quality of primary MMR studies, nor on the form in which these criteria should be grouped. None of the retrieved frameworks can yet truly be called a *critical appraisal tool*. What makes a CAF a good critical appraisal tool is the availability of a user guide, a clear scoring (numerical) or judging (narrative) system, the optimization of the tool through piloting it by intended end users (researchers doing a review including primary MMR evidence), and the availability of evidence regarding its psychometric properties.

Concerning the retrieved frameworks, we first of all notice that they often only enumerate and elaborate criteria that could be included in tools that critically appraise the methodological quality of primary MM studies, and are not (yet) modeled to the shape of a critical appraisal tool. Second, we notice that only some of these authors suggest a judging or grading frame for the methodological quality of the study that can assist the user (Alborz & McNally, 2004; Dybå et al., 2007; O'Cathain et al., 2008; Pluye et al., 2005, 2009a). Third, to our current knowledge none of the frameworks has yet been piloted by intended end users, apart from the authors that designed the frameworks.

An issue related to the user-friendliness of CAFs when conducting a systematic review including qualitative, quantitative, and MM primary research evidence concerns the original purpose of these frameworks: only 4 out of 13 frameworks state to be intentionally designed for use in systematic reviews (Alborz & McNally, 2004; Dybå et al., 2007; Pluye et al., 2005, 2009a). The authors of the other CAFs do mention the use of the developed frameworks for systematic reviewers, but do not state this to be the prior motive behind the development of the framework. Some other enumerated advantages of listing quality criteria for MMR are: answering to the increasing interest of commissioning and funding agencies in the quality of the research projects they fund (Bryman et al., 2008; O’Cathain et al., 2008), and more generally answering to the rising audit culture with a concern for judging quality of studies and research projects (Bryman et al., 2008; Caracelli & Riggin, 1994). Accordingly, it is suggested that the developed CAFs can not only serve as evaluation tools deciding on inclusion of primary studies in systematic reviews, but as well as evaluation tools for funding agency reviewers, for committees and advisors, for journal article reviewers, for practitioners intending to use the information described in the study, as well as for other researchers (Bryman et al., 2008; Creswell & Plano Clark, 2007; O’Cathain et al., 2008; Onwuegbuzie & Johnson, 2006) Additionally, Dellinger and Leech (2007) stress the use of the frameworks for organizing thoughts about a body of literature. Sale and Brazil (2004) mention the use of the CAFs for deducing a list of criteria which have to be reported in MM articles.

Recent developments and updates

For this paper, we systematically searched for manuscripts reporting on CAFs developed to evaluate the methodological quality of primary MMR articles published up to December 31, 2009. Many of these are currently being updated. For evaluating a MM study, Creswell and Plano Clark (2011, pp. 267-268) have recently suggested to use established standards for both the qualitative and quantitative study components, next to the following specific MM evaluation criteria: (a) collecting both quantitative and qualitative data, (b) employing persuasive and rigorous procedures in the methods of data collection and analysis, (c) integrating or mixing the two sources of data so that their combined use provides a better understanding of the research problem than one source or the other, (d) including the use of a MMR design and integrating all features of the study consistent with the design, (e) framing the study within philosophical assumptions, and (f) conveying the research using terms that are consistent with those being used in the MM field today. O’Cathain (2010) proposes a framework with 44 appraisal criteria situated within 8 domains of quality: planning quality,

design quality, data quality, interpretive rigor, inference transferability, reporting quality, synthesizability, and utility. Domains and items are based on the work of Caracelli and Riggan (1994), Creswell (2003), Creswell and Plano Clark (2007), Dellinger and Leech (2007), O’Cathain et al. (2008), Onwuegbuzie and Johnson (2006), Pluye et al. (2009a), and Teddlie and Tashakkori (2003, 2009), authors of which the frameworks were also included into our analysis. Thus combining criteria from many of the frameworks that were retrieved by our systematic search as well, O’Cathain (2010) presents a promising attempt towards a comprehensive CAF for primary MMR studies. However, as remarked by the author herself, the framework includes too many criteria, and some criteria tend to overlap. Addressing this drawback, O’Cathain and colleagues plan an international Delphi study in order to identify the *key* quality criteria for MMR within this extensive list. Such exercises hold the promise to arrive at a comprehensive tool for critically appraising primary MM studies.

Conclusion

The number of published primary MMR articles is growing steadily in several research domains, and researchers are challenged to include this type of studies in systematic reviews. However, a consensus on the critical appraisal of MM studies is lacking. By providing an overview of the available CAFs developed for the evaluation of the methodological quality of primary MMR articles as well as the criteria presented in these frameworks, we hope to contribute to the further development of guidance on the critical appraisal of primary MM studies that might potentially be included in systematic reviews. Developing a new CAF for primary MMR articles was not the aim of the present study. However, the list of identified criteria within the retrieved appraisal frameworks can be used to continue the dialogue on potential and necessary elements that have to be included in ready-to-use critical appraisal checklists for evaluating the methodological quality of primary MMR articles.

References

- * Alborz, A., & McNally, R. (2004). Developing methods for systematic reviewing in health services delivery and organization: An example from a review of access to health care for people with learning disabilities. Part 2. Evaluation of the literature - A practical guide. *Health Information and Libraries Journal*, 21, 227-236.
- Alise, M. A., & Teddlie, C. (2010). A continuation of the paradigm wars? Prevalence rates of methodological approaches across the social/behavioral sciences. *Journal of Mixed Methods Research*, 4, 103-126.
- Biesta, G. J. J. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 95-118). Thousand Oaks, CA: Sage.
- Bryman, A. (2006a). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology*, 9, 111-126.
- Bryman, A. (2006b). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6, 97-113.
- * Bryman, A., Becker, S., & Sempik, J. (2008). Quality criteria for quantitative, qualitative and mixed methods research: A view from social policy. *International Journal of Social Research Methodology*, 11, 261-276.
- * Caracelli, V., & Riggin, L. (1994). Mixed-method evaluation: Developing quality criteria through concept mapping. *Evaluation Practice*, 15, 139-152.
- Collins, K. M. T., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed-methods research in special education and beyond. *Journal of Mixed Methods Research*, 4, 103-126.
- Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). London, UK: Sage.
- Cooper, H. M., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Sage.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2008). *How mixed methods has developed*. Keynote address for the 4th Annual Mixed Methods Conference, Fitzwilliam College, Cambridge University, UK.
- Creswell, J. W. (2009). Mapping the field of mixed methods research. *Journal of Mixed Methods Research*, 3, 95-108.
- Creswell, J. W. (2010). Mapping the developing landscape of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 45-68). Thousand Oaks, CA: Sage.
- * Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Day, C., Sammons, P., & Gu, Q. (2008). Combining qualitative and quantitative methodologies in research on teachers' lives, work, and effectiveness: From integration to synergy. *Educational Researcher*, 37, 330-342.
- Dellinger, A. B. (2005). Validity and the review of literature. *Research in the Schools*, 12(2), 41-54.
- * Dellinger, A. B., & Leech, N. L. (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, 1, 359-375.

- * Dybå, T., Dingsøy, T., & Hanssen, G. K. (2007). Applying systematic reviews to diverse study types: An experience report. *Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement (ESEM'07)*, IEEE Computer Society, 225-234.
- Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4, 6-16.
- Fidel, R. (2008). Are we there yet? Mixed methods research in library and information science. *Library & Information Science Research*, 30, 265-272.
- Gaber, J. (2000). Meta-needs assessment. *Evaluation and Program Planning*, 23, 139-147.
- Gliner, J. A., & Morgan, G. A. (2000). *Research methods in applied settings: An integrated approach to design and analysis*. Mahwah, NJ: Lawrence Erlbaum.
- * Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey Bass.
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2, 7-22.
- Greene, J. C., Benjamin, L., & Goodyear, L. (2001). The merits of mixing methods in evaluation. *Evaluation*, 7, 25-44.
- Greene, J. C., & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 91-110). Thousand Oaks, CA: Sage.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274.
- Greene, J. C., & Hall, J. N. (2010). Dialectics and pragmatism: Being of consequence. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 119-143). Thousand Oaks, CA: Sage.
- Haahr, M. (1998-2011). RANDOM.ORG. Retrieved from <http://www.random.org/>
- Hall, B., & Howard, K. (2008). A synergistic approach - Conducting mixed methods research with typological and systemic design consideration. *Journal of Mixed Methods Research*, 2, 248-269.
- Hannes, K., & Lockwood, C. (2011). Pragmatism as the philosophical underpinning of the Joanna Briggs meta-aggregative approach to qualitative evidence synthesis. *Journal of Advanced Nursing*, 67, 1632-1642.
- Harden, A., & Thomas, J. (2005). Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology*, 8, 257-271.
- Harden, A., & Thomas, J. (2010). Mixed methods and systematic reviews: Examples and emerging issues. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 749-774). Thousand Oaks, CA: Sage.
- Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2009). An examination of research methods in mathematics education (1995-2005). *Journal of Mixed Methods Research*, 3, 26-41.
- Heyvaert, M., Maes, B., & Onghena, P. (2011a). Applying mixed methods research at the synthesis level: An overview. *Research in the Schools*, 18(1), 12-24.
- Heyvaert, M., Maes, B., & Onghena, P. (2011b). Mixed methods research synthesis: Definition, framework, and potential. *Quality & Quantity* (published online: 9 July 2011).
- Higgins, J. P. T., Altman, D. G., the Cochrane Statistical Methods Group, & the Cochrane Bias Methods Group (2008). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 5.0.1). Retrieved from <http://www.cochrane-handbook.org>
- Higgins, J. P. T., & Green, S. (Eds.). (2009). *Cochrane handbook for systematic reviews of interventions* (Version 5.0.2). Retrieved from <http://www.cochrane-handbook.org>
- Hurmerinta-Peltomaki, L., & Nummela, N. (2006). Mixed methods in international business research: A value-added perspective. *Management International Review*, 46, 439-459.

- Hutchinson, S. R., & Lovell, C. D. (2004). A review of methodological characteristics of research published in key journals in higher education: Implications for graduate research teaching. *Research in Higher Education, 45*, 383-403.
- International Centre for Allied Health Evidence, University of South Australia. *Critical appraisal tools*. Retrieved from <http://www.unisa.edu.au/cahe/Resources/CAT/default.asp>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*, 14-26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*, 112-133.
- Katrak, P., Bialocerkowski, A. E., Massy-Westropp, N., Kumar, V. S. S., & Grimmer, K. A. (2004). A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology, 4*, 22-33.
- Khan, K. S., Riet, G., Popay, J., Nixon, J., & Kleijnen, J. (2001). Undertaking systematic reviews of research on effectiveness. *Centre for Reviews and Dissemination Report, 4*, 1-20.
- ** Leech, N. L., Dellinger, A. B., Brannagan, K. B., & Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research, 4*, 17-31.
- Letts, L., Wilkins, S., Law, M., Stewart, D., Bosch, J., & Westmorland, M. (2007). *Critical Review Form – Qualitative Studies* (Version 2.0). Retrieved from http://www.srs-mcmaster.ca/Portals/20/pdf/ebp/qualreview_version2.0.pdf
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.
- Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In: D. D. Williams (ed.), *Naturalistic evaluation. New directions for program evaluation, No. 90* (pp. 78-84). San Francisco, CA: Jossey Bass.
- Major, C. H., & Savin-Baden, M. (2010). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*. London, UK: Routledge.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*, 279-300.
- Mertens, D. M. (2010). Transformative mixed methods research. *Qualitative Inquiry, 16*, 469-474.
- Mertens, D. M. (2012). Transformative mixed methods: Addressing inequities. *American Behavioral Scientist, 56*, 802-813.
- Mertens, D. M., Bledsoe, K., Sullivan, M., & Wilson, A. (2010). Utilization of mixed methods for transformative purposes. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 193-214). Thousand Oaks, CA: Sage.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.
- Moffatt, S., White, M., Mackintosh, J., & Howel, D. (2006). Using quantitative and qualitative data in health services research - What happens when mixed method findings conflict? *BMC Health Services Research, 6*, 28-38.
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*, 48-76.
- Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale: Southern Illinois University Press.
- O'Cathain, A. (2010). Assessing the quality of mixed methods research: Towards a comprehensive framework. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (2nd ed., pp. 531-555). Thousand Oaks, CA: Sage.
- * O'Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research and Policy, 13*, 92-98.

Onwuegbuzie, A. J., Collins, K. M. T., Leech, N. L., Dellinger, A. B., & Jiao, Q. G. (2010). A meta-framework for conducting mixed research syntheses for stress and coping and beyond. In G. S. Gates, W. H. Gmelch, & M. Wolverson (Series Eds.) & K. M. T. Collins, A. J. Onwuegbuzie, & Q. G. Jiao (Vol. Eds.), *Toward a broader understanding of stress and coping: Mixed methods approaches* (pp. 169-212). The Research on Stress and Coping in Education Series: Vol. 5. Charlotte, NC: Information Age Publishing.

* Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13*, 48-63.

** Onwuegbuzie, A. J., Johnson, R. B., & Collins, K. M. T. (2011). Assessing legitimization in mixed research: A new framework. *Quality & Quantity, 45*, 1253-1271 .

Onwuegbuzie, A. J., & Leech, N. L. (2005a). Taking the "Q" out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quality & Quantity, 39*, 267-296.

Onwuegbuzie, A. J., & Leech, N. L. (2005b). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology: Theory and Practice, 8*, 375-387.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review - A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(Suppl.1), 21-34.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell Publishing.

* Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009a, November). *A proposal for concomitantly appraising qualitative, quantitative and mixed methods primary studies included in systematic mixed studies reviews*. Paper presented at the 37th Annual Meeting of the North American Primary Care Research Group (NAPCRG), Montreal, Canada.

** Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009b). 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*, 529-546.

* 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*, 745-768.

Public Health Resource Unit (2006a). *Critical Appraisal Skills Programme (CASP). 10 questions to help you make sense of qualitative research*. Retrieved from <http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/Qualitative%20Appraisal%20Tool.pdf>

Public Health Resource Unit (2006b). *Critical Appraisal Skills Programme (CASP). 11 questions to help you make sense of a case control study*. Retrieved from <http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/Case%20Control%2011%20Questions.pdf>

Saini, M., & Shlonsky, A. (2012). *Systematic synthesis of qualitative research*. Oxford: Oxford University Press.

* Sale, J. E. M., & Brazil, K. (2004). A strategy to identify critical appraisal criteria for primary mixed-method studies. *Quality & Quantity, 38*, 351-365.

Sandelowski, M., Voils, C. I., & Barroso, J. (2006). Defining and designing mixed research synthesis studies. *Research in the Schools, 13*(1), 29-40.

Sweetman, D., Badiie, M., & Creswell, J. W. (2010). Use of the transformative framework in mixed methods studies. *Qualitative Inquiry, 16*, 441-454.

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research, 1*, 3-7.

- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003a). *Handbook of mixed methods in social and behavioural research*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (2003b). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671-701). Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (2006, April). *Validity issues in mixed methods research: Calling for an integrative framework*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- ** Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.), *Advances in mixed methods research: Theories and applications* (pp. 101-119). London, UK: Sage.
- ** Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 3-50). Thousand Oaks, CA: Sage.
- * Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
- Truscott, D., Swars, S., Smith, S., Thornton-Reid, F., Zhao, Y., & Dooley, C. (2010). A cross-disciplinary examination of the prevalence of mixed methods in educational research: 1995-2005. *International Journal of Social Research Methodology*, 13, 317-328.
- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11(4), 522-537.
- Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52, 546-553.
- Young, J. M., & Solomon, M. J. (2009). How to critically appraise an article. *Nature Clinical Practice Gastroenterology & Hepatology*, 6, 82-91.

* References to studies included in this review.

** References to studies excluded from this review, because they are manuscripts that describe the same CAF as presented in one of the already included studies. In each case, we choose to include the most comprehensive CAF. In case where the separate manuscripts contained exactly the same framework, we included the original manuscript.

Table 1

Overview of the criteria included in the 13 retrieved critical appraisal frameworks developed to evaluate primary mixed methods research articles, with the frequency that each criterion is mentioned in the frameworks

Critical appraisal criteria		Frequency
Criteria to score separately the methodological quality of the qualitative and quantitative strands of a study	Criteria for scoring the qualitative strand of the study	9
	Criteria for scoring the quantitative strand of the study	9
Criteria that are explicitly concerned with mixed methods research	Mixing and integration of methods	9
	Rationale for mixing methods stated	4
Generic critical appraisal criteria	Design	9
	Interpretation, conclusions, inferences, and implications	8
	Data analysis	7
	Research aims and questions	5
	Sampling and data collection	4
	Theoretical framework	3
	Impact of investigator	3
	Transparency	2
Context	2	