Assessment of transparency, structure, and coherence: A systematic review of empirical research articles on social movements in Latin America

February, 2014
Minor thesis by Sven da Silva (900816-762-040)
Research Master Variant International Development Studies
Supervised by Dr P.A Tamás and Dr J.K Kampen
Thesis code: YRM-80324
“It’s a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty — a kind of leaning over backwards”.

“An issue can certainly be viewed from multiple perspectives, but anybody who claims that justifies misrepresenting information or denies the value of truth and objective analysis is really bullshitting.”
___(Litman, 2012, p. 3 using the work of Harry Frankfurt)
# Table of contents

Acknowledgements ........................................................................................................................................................... 5  
Summary ............................................................................................................................................................................. 6  
1. Introduction ................................................................................................................................................................. 12  
2. The Framework ............................................................................................................................................................ 15  
   2.1 “Systematic” quality assessment instruments ........................................................................................................... 15  
   2.2 Operationalisation of transparency .......................................................................................................................... 19  
   2.2 Operationalisation of structuredness ........................................................................................................................ 21  
   2.3 Operationalisation of coherence ............................................................................................................................... 22  
3. Data extraction form .................................................................................................................................................... 24  
   3.1 Development ........................................................................................................................................................... 24  
   3.2 Application ............................................................................................................................................................. 29  
   3.3 Efficiency, reliability and internal validity .................................................................................................................. 32  
4. Sample ........................................................................................................................................................................... 34  
   4.1 Search syntax ............................................................................................................................................................ 34  
   4.2 Classification .......................................................................................................................................................... 35  
   4.3 Systematic random sampling ................................................................................................................................... 36  
5. Results ............................................................................................................................................................................ 38  
   5.1 The sample .............................................................................................................................................................. 38  
   5.2 Transparency ......................................................................................................................................................... 41  
   5.3 Structure ............................................................................................................................................................... 45  
   5.4 Coherence .............................................................................................................................................................. 47  
6. Discussion ...................................................................................................................................................................... 50  
7. Limitations .................................................................................................................................................................... 53  
8. Conclusion and Outlook ............................................................................................................................................... 55  
References ......................................................................................................................................................................... 56  
Appendix 1: Data extraction form ................................................................................................................................... 59  
Appendix 2: Articles included in the review (N=64) ...................................................................................................... 61  
Appendix 3: Articles classified as secondary research (N=93) .................................................................................. 64  
Appendix 4: Web Of Science result (N=549) .................................................................................................................. 68
Acknowledgements

The author thanks the supervisors of this thesis for the amount of time this thesis asked from them. He also thanks the organisers and reviewers of the course 'from thesis to journal manuscript' for their comments during first drafts of this thesis.
Summary

1. Introduction

Social movement researchers and policymakers consume research on social movements in Latin America. Journals such as the Journal of Peasant Studies publish social movement research articles which are consumed by development planners. These development planners need useful knowledge and should be able to make an informed decision on inferences made in the articles they consume. While useful knowledge is communicated through adequate reporting, little is known about reporting practices in the area of social movement research. A systematic review would be useful for making judgements about reporting practices (Petticrew & Roberts, 2006). To the knowledge of the author, there have not been systematic reviews conducted before in the topic area of social movement research.

Therefore, empirical research articles on social movements in Latin America were taken as a case for a systematic review using a data extraction form based on a checklist developed by Kampen & Tamás (2013) intended for consumers of policy supporting research to screen articles for fatal flaws. Their checklist operates on the level of internal coherence of research reports rather than the level of research quality. However – together with similar checklists (Carroll, Booth, & Lloyd-Jones, 2012; M. Dixon-Woods, et al., 2006b; Mary Dixon-Woods, Shaw, Agarwal, & Smith, 2004) – their checklist implicitly assumed that articles would supply all relevant information regarding the research process such as the research question, data collection and analysis etc (transparency) in a form which would enable consumers of research articles to separately assess aspects of the empirical research process (structure).

During the transformation of the checklist into a data extraction form the problem was encountered of having to apply the checklist to articles for which these assumptions did not hold. The consequence being that it proved difficult to systematically assess internal coherence in case articles were non-transparent or unstructured. Other evaluation frameworks and checklists did not provide useful help since they also relied on the implicit assumptions in Kampen & Tamás (2013). Furthermore, these frameworks and checklist tended to (1) assume evaluators of research articles to have a considerable amount of knowledge on research methodology or (2) they provided non-quantitative quality indicators. The latter is according to the author why “systematic” in systematic quality evaluation is not solid enough yet, especially when the review includes qualitative research (Mary Dixon-Woods, et al., 2006a; Margarete Sandelowski, 2008).

Thus, it was necessary to assess first whether it would be possible to assess questions regarding internal coherence. That is, this systematic review asked: to what extent are empirical social science articles on Latin American social movements transparent and structured? And to what extent are transparent articles also coherent?

A transparent article would supply the following seven aspects: (1) clear central research question(s); (2) clear data collection method(s); (3) a description of the sampling strategy; (4) a sample size for each data collection method; (5) a description of the analysis method; (6) clear conclusion(s), and (7) a reflexive account. A structured article would imply transparency, but in a structured article (1) the body text supplied a research question, data collection method and conclusion; (2) with a possibility to link research questions with the data collection methods, and conclusions; (3) and a possibility to distinguish theory from data, data from results, and results from conclusions. The last mentioned would be facilitated by (3a) the
following sequence of locations in an article; central research question first, data collection methods second and conclusions third, together with (3b) considerable proportion of the document of applied methods in the article (data collection, sampling and analysis).

In a coherent article the research questions would match the methods and conclusions. In a match (1) conclusions would answer central research questions; (2) the ‘object of interest’ would be the same as the ‘object of research’ and ‘object of conclusion’; (3) the data collection methods would be appropriate for the highest reach of the conclusions; (4) the study design would be appropriate for the reach of the conclusions; (5) the study design would be appropriate for the type of research questions; (6) the highest reach of the conclusions would match the highest reach of the research questions; (7) the types of questions would be similar to the types of conclusions.

2. Method

To explore reporting practices in Latin American social movement research, a systematic review was conducted on the structure and content of accessible English and Spanish primary research articles on Latin American social movements. A systematic review would provide a replicable and transparent procedure. The review involved (1) the development and test of a data extraction form, (2) creating a population of Latin American social movement articles, (3) data extraction using Atlas.ti, a qualitative data analysis software program and (4) data analysis consisting mainly of frequency counts and cross-tabs in SPSS. On request it is possible to access the list of all search result papers together with a classification made based on population criteria; data from Atlas.ti; and SPSS data and syntax.

The data extraction form was developed in a survey form based on a checklist developed by Tamás and Kampen (2013). The data extraction for included both questions on an article level (for example study design) and questions for each individual research question, data collection method, and conclusion (for example types). In the software program Atlas.ti, answers to survey questions were tagged – where possible – to corresponding locations in articles.

The population was created by a developed search syntax which was entered, on January 16th 2013, into ISI Web of Science. The search resulted in 549 articles of which the abstracts were scanned. To be included, a paper had to: be available through the Wageningen University Library; be on social movement(s) located in Latin America; primary social science research; and in the English or Spanish language. In total 219 articles were available. Articles were systematically random sampled in cohorts of ten and the complete data extraction form was applied to 64 articles. Data extraction proceeded in the following order for each cohort: (1) all ten articles on research questions; (2) all ten articles on data collection method; (3) all ten articles on conclusion and implications; and finally (4) all ten articles on study design and the degree of structuredness. The strategy to extract data in steps per cohort rather than complete data extraction per article would strengthen internal validity. The data extracted was in most cases collapsed in order to generate variables – on an article level – with the values missing, unclear, or clear.

Simple indexes were developed for transparency and structuredness. The score on the transparency index was calculated by ((the number of clearly present aspects out of seven multiplied by two + the total number of unclearly present aspects out of seven multiplied by one)/ (14, the total score in case all seven aspects would be clearly present). For structuredness, articles were rated on a three-point scale ranging from unstructured towards moderately
structured towards structured. Articles which were to some extent transparent were assessed for seven possible incoherencies.

3. Results

Major results regarding the sample were:

- There has been an increasing trend in number of publications of Latin American social movement research articles, with most articles published between 2007 and 2010.
- Most articles (8 out of 64) were published in the journal *Latin American Perspectives*.
- Articles in the population had a mean of 0.57 citations per year (with standard deviation 0.85). Articles included in this systematic review had a mean of 0.54 citations per year (with standard deviation 0.60).

Major results regarding transparency were:

- One-third of the articles (23 out of 64) were not transparent regarding the research questions. This includes three articles which did not supply a central research question and 20 articles with unclear research questions.
- Two-thirds of the articles (43 out of 64) were not transparent regarding the data collection methods used. This includes 12 articles which did not explicitly supply a data collection method and 31 articles with unclear data collection methods.
- 83% of the articles (53 out of 64) dedicated no attention at all to the sampling strategy (random, non-random) used.
- Half of the articles (31 out of 64) did not report any sample size and almost half of the articles (30 out of 64) had incompletely reported sample sizes.
- More than half of the articles (36 out of 64) dedicated no attention at all to the analysis method used.
- Almost two-thirds of the articles (41 out of 64) were not transparent regarding the conclusions reached. This includes six articles which either did not supply a CRQ or missed a link to the CRQ and 35 articles which supplied unclear answers to the CRQ.
- Two-thirds of the articles (43 out of 64) did not supply a reflexive account. This includes 30 articles which neither mentioned of fallacies nor mentioned of instrument effects, and 13 articles which supplied either an unclear mention of fallacies or instrument effect.
- Quite often articles had less than four out of seven aspects of transparency clearly present. The score on the transparency index ranged from 0.07 towards 0.93, with a mean of 0.46 and standard deviation of 0.23.

Major results regarding structuredness were:

- Two-thirds of the articles (42 out of 64) were unstructured and almost one-third of the articles (19 out of 64) were moderately structured.
- One-third of the articles (21 out of 64) did not supply research questions or data collection methods in the body text. Quite often the first shown location of the data collection methods was footnotes (13 articles out of 64). In eight articles a data collection was neither supplied in the body text, nor in footnotes. In two articles only the abstract supplied a research question.
• A quarter of the articles (16 out of 64) followed the sequence of research question first, the data collection method second and conclusion third.

• Almost half of the articles (30 out of 64) had little influence of data gathered on the conclusions reached. The same amount of articles had conclusions based both on the data gathered and theory.

• In almost half of the articles (29 out of 64) the proportion of document relating to methods used (sum of data collection, sampling and analysis) was lower than one percent.

Related to coherence:

• A small number of articles remained for the assessment of coherence. One article clearly supplied five aspects for an assessment of coherence. For the assessment of coherence non-transparent articles regarding aspects being assessed were excluded.

• Two-thirds of the articles which had both clear central research questions and clear conclusions (11 out of 17) had an equal number or central research questions and conclusions which answered a central research question.

• None of the articles were completely coherent regarding to the object of interest, object of research, and object of conclusion. From five articles which had clear research questions, data collection methods and conclusions, three articles were completely incoherent, one was for 25% coherent and one was for 75% coherent. Overall, for all 64 articles, the minimum was 0% and the maximum was 75%, with a mean of 22% and a standard deviation equal to 19%.

• One out of seven articles with both clear data collection methods and conclusions had a mismatch in the reporting of the data collection method and the highest reach of the conclusions.

• Two out of 23 articles with clear conclusions had a mismatch between the highest reach of conclusions and the study design.

• A quarter of the articles with clear research questions (10 out of 41 articles) were incoherent regarding the study design and the type of research questions.

• One article out of 17 articles with both clear research questions and conclusions had a mismatch between the highest reach of research questions and the highest reach of conclusions.

• Seven articles with both clear and an equal number of research questions and conclusions were all coherent regarding the types of research questions and the types of conclusions.

4. Discussion

The initial plan of this systematic review was to subject empirical research articles on social movements in Latin America to a BS checklist (Kampen & Tamás, 2013) which would provide insights into the reporting coherence of these articles. However, it was not anticipated beforehand that these articles would severely lack in transparency and structure such that systematic application of a BS checklist would not be systematic. Applying the checklist to non-transparent and unstructured articles would imply judgements to be made based on inferences about the research as executed rather than systematically based on research as reported.
As such, this review took a step back during the development of the data extraction form to extract data regarding transparency and structuredness of research articles on social movements in Latin America. The approach was not based on strict criteria for reporting methods, and unclear reporting was also taken into account for the assessment. Since the assessment of coherence of research articles depended on whether articles were transparent and structured, the results of this systematic review regarding coherence of research articles should not be considered as rigorous results.

This systematic review suggests a change of use of the BS checklist; *a quick scan through articles for transparency and structure before application of the BS checklist.* This change would be necessary since non-presence, or unclear presence, of one of the following makes systematic quality assessments unsystematic; (a) central research questions (b) data collection- (c) sampling- and (d) analysis- methods; (e) study design; (f) results; (g) theory or (h) conclusions. This would be a simple scan involving a count of the mention of a number of determinants of (reporting) quality.

At the moment different checklists mix questions of transparency and adequacy and include a possibility for evaluators to add written comments for questions in the checklist. In this form, a final decision on quality remains non-systematic in the sense that step-by-step guidelines for evaluators and standard measurement for specific items in checklists are missing. Rating articles for transparency, such as proposed by this systematic review, could make assessments systematic because (1) by using a rating system different studies are handled the same way based on interpretability (2) differences between evaluators of research could easily be identified and (3) a minimum total score could be set as standard for inclusion in policy supporting systematic reviews.

As a final note; it is important to be careful in calling BS on research articles. It may be questioned whether the articles in the population – taking the type of research into account – should have been subjected to a “checklist style” bullshit detection in the first place. However, there is no reason why qualitative (social movement) research cannot ensure rigor (Krefting, 1991; Popay & Williams, 1998; Seale & Silverman, 1997) and consumers of research articles need to be able to see how the rigor has been achieved before they can accept the findings of a study. Rather than being a threat to qualitative research, the BS checklist developed by Tamás and Kampen (2013), and the transparency index from the current thesis, is a caution against other threats to qualitative research, such as the development of real world descriptive theories based on an over-reliance of (cherry-picked) verbal representations. In order to assess how theories emerge, articles should first be transparent and structured.

5. Conclusion and Outlook

The first question asked was: *to what extent are empirical social science articles on Latin American social movements transparent and structured?* Latin American social movement articles included in this systematic review were neither transparent nor structured sufficiently. The articles were transparent to the extent that it was possible to identify research questions, data collection methods and conclusions. However, these three aspects were quite often unclear. Analysis methods, sampling methods, and a reflexive account were often not supplied at all. Articles did not go further regarding structure then the overlap with transparency, i.e. the possibility to identify research questions, data collection methods, and conclusions. Making a link between these aspects, together with a separation between the research process and the research results was in the majority of cases not possible.
The second question asked was: to what extent are transparent articles also coherent? A reliable conclusion on the coherence of research articles could not be made since a small number of articles remained after excluding non-transparent articles for the seven individual assessments. However, there is reason for concern regarding (1) incoherence between the study designs and the type of research questions asked, and (2) incoherence between the number of conclusions and number of central research questions.

It is proposed that (Latin American) social movement researchers use available guidelines for reporting their research, such as COREQ (Tong, Sainsbury, & Craig, 2007) or an older guide developed by Knafl and Howard (1984). For editors and reviewers of research it is proposed that they use indicators such as the developed transparency index in order to publish transparent articles or include adequately reported transparent articles in reviews.

**Keywords:** Latin America; social movements; systematic reviews; transparency; structure; coherence; reporting quality
1. Introduction

Social movement researchers and policymakers consume research on social movements in Latin America. Social movement researchers often compare the movement they study with other movements. For comparisons they are dependent on research articles asking similar questions and applying similar methods. These researchers should be able to judge whether the research articles they consume fit the purpose of comparison. Ideally they should also be able to judge whether the articles they consume were based on solid research, since a comparison based on flawed research can also be expected to be flawed. Policy makers such as development planners also consume social movement articles published in journals such as the *Journal of Peasant Studies*. These development planners need useful knowledge. They should be able to make an informed decision on inferences made in the articles they consume. Both the social movement researcher and the policymaker should be able to observe how the articles they consume link collected data with theory. Since it is not efficient for them to contact authors of published research for relevant methodological aspects missing in the publication, they are limited by what is reported in a publication for a decision on the usefulness of the knowledge.

Although useful knowledge is communicated through adequate reporting, there is little known about reporting practices in the area of social movement research. That we know little about reporting practices is probably influenced by a lack of discussion on research methods in the field of social movement research (B. Klandermans & Staggenborg, 2002, p. vii), together with an inability to generate quality standards for systematic quality assessment of qualitative research (Mary Dixon-Woods, et al., 2004; N. Mays & Pope, 2000; Seale, 1999).

However, a useful starting point for an assessment of reporting practices was an article by Kampen and Tamás (2013) who developed a simple “BS checklist” intended for consumers of policy supporting research to quickly identify fatal flaws in research articles. It was a useful starting point, because (1) the checklist bases evaluation of research articles with diverse designs on the research questions posed rather than on methodological or epistemological preferences and (2) it assesses coherence of research articles rather than quality of research. Their “BS checklist” is not methodologically sophisticated since it is not concerned with existing quality criteria to judge the data obtained from specific data collection procedures, nor is it concerned with quality criteria developed to judge qualitative analyses of data.

The initial strategy of this thesis was to develop a data extraction form based on their “BS checklist” for a systematic review of research articles on social movements in Latin America. A systematic review would be useful to explore reporting practices (Petticrew & Roberts, 2006, p. 21) of research on social movements in Latin America. To the knowledge of the author, a systematic review has not been conducted before in this research area. The central research question initially was: is there a “mismatch” between questions asked, methods used and

---

1 It is remarkable that systematic assessment of methodological rigor in social movement research has not been attempted before. There have been calls for more synthesis in the field of social movement research by McAdam, Tilly and Tarrow (2001). They use the term “contentious politics” to combine different kinds of uprisings hoping to find commonalities among them. It can be assumed that in order to make valid comparisons based on empirical research, clear reporting of the consumed research is necessary.

2 Checklists and quality assessments are often criticized for being “methodological fundamentalist” (Norman K Denzin, 2009; Norman K. Denzin, Lincoln, & Giardina, 2011; House, 2008), i.e. where the randomized control trial is seen as the best method. Also, according to Reid and Gough (2000, p. 60) “criteria for judging research quality contain within them, implicitly or explicitly, a defining view of what research is, and perhaps what it should be”.

3 There is a growing interest in systematic reviews, because they have an ability to provide a solid evidence base to inform policymakers (Mary Dixon-Woods, et al., 2007; Gough, Oliver, & Thomas, 2012; Petticrew & Roberts, 2006).
conclusions reached in the reporting of empirical social science research on Latin American social movements?

During test rounds in the development phase of the data extraction form it was noticed that the "BS checklist" was based on two assumptions; (1) that articles were transparent, meaning that articles would supply all relevant information regarding the research process such as the research question, data collection, analysis, sampling methods, results etc and (2) that articles would be structured, meaning that it would be possible to quickly locate sections such as "methods", "analysis" and "limitations" such that it would enable consumers of research articles to separately assess aspects of the empirical research process. The latter would thus facilitate making a distinction between research question, methods, theory, results, conclusions and implications.

However, the design and purpose of the data extraction form had to be adapted because it turned out these assumptions did not hold. Before assessing whether a mismatch was present between two aspects it was necessary to detect whether articles supplied the necessary individual aspects. If it was not possible to detect research questions, data collection methods, sampling methods, analysis methods, and conclusions, a statement about internal coherence would prove difficult.

The central research question was reformulated towards: to what extent are empirical social science articles on Latin American social movements transparent and structured? And to what extent are transparent articles also coherent? In order to answer this question, simple indexes were developed for transparency and structuredness. For coherence it was asse whether some often found incoherencies could be detected in articles which were partially transparent.

A transparent article would supply the following seven aspects: (1) clear central research question(s); (2) clear data collection method(s); (3) the sampling process; (4) sample size(s); (5) the analysis method; (6) clear conclusions; (7) a reflexive account.

A structured article would imply transparency, but in a structured article: (1) the body text would supply a research question, data collection method, and conclusion (2) with a possibility to link these three aspects and (3) and enable making a distinction between theory and data, data and results, and between results and conclusions. The last mentioned would be facilitated by (3a) supplying central research question first, data collection methods second, and conclusions third, together with (3b) a considerable proportion of the article documenting the applied methods (data collection, sampling and analysis).

Criteria for a coherent article also minimally overlap with criteria for structuredness. However, for in a coherent article there would be a match between the research questions, data collection- and analysis methods and conclusions. In a match (1) conclusions would answer central research questions; (2) the ‘object of interest’ would be the same as the ‘object of research’ and ‘object of conclusion’; (3) the data collection methods would be appropriate for the highest reach of the conclusion; (4) the study design would be appropriate for the highest reach of the conclusions; (5) the study design would be appropriate for the type of research questions; (6) the highest reach of the conclusions would be appropriate for the highest reach of the research questions and (7) the type of questions asked would be similar to the type of conclusions reached.

The following chapter positions the focus on transparency, structure and coherence in literature concerned with “quality” in qualitative research. The chapter presents the operationalisation of these concepts and justifies the operationalisation based on a comparison
with other frameworks and checklists used in the quality assessment phase of systematic reviews. The third chapter describes the development and use of the data extraction form, together with an assessment of its efficiency, reliability and internal validity. The chapter describes major difficulties encountered during the development phase and how they were dealt with. The fourth chapter describes the search and sampling strategy. The fifth chapter presents the results and chapter six discusses these results and suggests a change in the use of the "BS checklist" by Kampen and Tamás. Chapter seven presents limits of this systematic review and the last chapter provides a short answer to the central research question and makes suggestions for editors, reviews and social movement researchers to improve reporting of research articles on social movements in Latin America.
2. The Framework

2.1 “Systematic” quality assessment instruments

Systematic reviews focus on a research question and then follow the steps of identification, appraisal, selection, and synthesis of research evidence related to that question (Gough, et al., 2012; Petticrew & Roberts, 2006). Although systematic reviews of randomized control trials are highly influential in evidence-based medicine, they are emerging in the social sciences. Increasing value is given to them to support policy and practice (Mary Dixon-Woods, et al., 2007). Quality assessments of individual studies are a crucial part of the systematic review process since such assessments improve the quality and credibility of the review itself (Gough, et al., 2012; Petticrew & Roberts, 2006).

While quality assessments are essential, there remains dispute about inclusion of qualitative research in systematic reviews because “quality” in qualitative research is highly contested (Mary Dixon-Woods, et al., 2004; Frances Ryan, 2007; N. Mays & Pope, 2000; Seale, 1999; Spencer, Ritchie, Lewis, & Dillon, 2003; Walsh & Downe, 2006). There is no consensus for the application of concepts such as bias, validity, and reliability to qualitative research (Ambert, Adler, Adler, & Detzner, 1995; Hannes, Lockwood, & Pearson, 2010; Rolfe, 2006). For example, the criterion of multiple coding for internal validity is meaningless for those who want to show the existence of “multiple realities”. In the “quality in qualitative research” debate there are a number of positions (discussion in Rolfe, 2006). Some argue that all research should be evaluated on the same criteria of scientific validity. Others argue that qualitative research should be evaluated on “its own merits” (Spencer, et al., 2003). A last position is taken up by those who argue that each individual study should be individually appraised\(^5\) taking aesthetic and rhetorical concerns into account (M. Sandelowski & Barroso, 2002). As such, no quality assessment checklist fits both types of research.

However, the “BS checklist” by Kampen & Tamás (2013) assesses coherence of reports\(^4\)

\(^{4}\) While there is no consensus on quality in qualitative research, there are frameworks for rigour in qualitative work (Nicholas Mays & Pope, 1995; Seale & Silverman, 1997; Tracy, 2010). Guba’s constructs have won considerable favour (Shenton, 2004). Guba’s constructs, correspond to criteria employed by the positivist investigator, although he distances himself from the positivism:

- Credibility (in preference to internal validity): a true picture of the phenomenon under scrutiny is being presented.
- Transferability (in preference to external validity/generalizability: provide sufficient detail of the context of the fieldwork for a reader to be able to decide whether the prevailing environment is similar to another situation).
- Dependability (in preference to reliability): researchers should at least strive to enable a future investigator to repeat the study.
- Confirmability (in preference to objectivity): findings are because of data analysis and not researchers’ predispositions.

However, a number of checklists for quality in qualitative research have been criticized for not assuring the quality of research (for example Rolfe, 2006); not taking different epistemological assumptions and unique methods into account (Anastas, 2004) and seen as a danger leading to a situation where “the tail (the checklist) is waging the dog (the qualitative research)” (Barbour, 2001).

\(^{5}\) Sandelowski and Barroso for example state (2002, p. 78) that they “prefer the word appraisal as opposed to evaluation, as appraisal more explicitly encompasses understanding in addition to estimating value”. According to them, a research report “must be understood, or appreciated, for what it is before it can be judged as a good or bad example of its kind. Appreciation here means the exercise of wise judgment and keen insight in recognizing the nature and merits of a work”. They then state that “if writers of qualitative research reports have an obligation to write well, readers of these reports have an obligation to read well” (ibid, p. 78).
rather than quality of research, since research reports are often the only site for consumers of research to evaluate the research they consume. Rather than provide detailed methodological descriptions for a quality assessment their decision protocol checklist suggested a number of simple questions to detect fatal flaws in research articles. Similarly, Dixon-Woods et al (2006b) preferred to use five prompts to assess the reporting of the research as an initial step in quality assessment for a systematic review which included qualitative research. However, their prompts did not include a users’ guide and in the end their prompts were not prioritized above the potential relevance of studies. A small number of papers were excluded because they judged inadequately reported articles as being potentially of high relevance.

During the transformation of the “BS checklist” into a data extraction form it proved difficult to systematically assess – i.e. quantify step-by-step – internal coherence in case articles were not transparent or structured. It turned out that many quality assessment instruments, the “BS checklist” included, do not make a practical distinction between reporting adequacy (auditability/transparency) and procedural appropriateness (validity/credibility) (Mary Dixon-Woods, et al., 2004; Nicholas Mays & Pope, 1995; Petticrew & Roberts, 2006; M. Sandelowski & Barroso, 2002).

A number of frameworks mix questions of transparency with questions of coherence or adequacy. For example the question “Are sampling, data collection and analysis appropriate to the research question?” (Mary Dixon-Woods, et al., 2004) rightly requires first to answer “Are sampling, data collection and analysis are clearly described in the articles?”. However, answering the latter question in turn assumes that it is possible to detect sampling, data collection, and analysis methods in the article. The prompts of Dixon-Woods et al. (2006b; 2004) are not specific in case it is found that an article either did not report sampling, data collection or analysis.

Other frameworks point to transparency but "act as if" it is possible to continue an evaluation of research quality. For example, the lengthy quality assessment framework for qualitative research by Spencer et al. (2003) has transparency, auditability, and reporting at the end of the framework. Also, Pawson et al. (2003, p. ix) have developed a framework used in systematic reviews (Gough, et al., 2012), with the acronym TAPUPAS. Transparency is positioned first in the framework. No mention is made however that judgement about other aspects (Accuracy, Purposivity, Utility, Propriety, Accessibility, and Specificity) is dependent on transparency.

In short, it turned out that a number of instruments were not suitable for the purpose of this systematic review, i.e. quantify coherence for a number of research articles. The data extraction form had to be adapted and used differently, since, as Carroll et al. (2012) have also pointed out, it is not possible to assess the validity, credibility, or trustworthiness, if authors do

---

6 Sandelowski and Barroso (2002) also took a similar approach, including inadequately reported articles in case they were judged to be of high relevance. The dilemma in deciding between the potential impact of findings and the quality of the research execution or reporting practice was also shown in a study which used three methods for appraising qualitative research to compare how reviewers categorized candidate papers for inclusion in a systematic review showed (Mary Dixon-Woods, et al., 2007). The paper of that study ended with a warning that (2007, p. 47) "those conducting systematic reviews that include qualitative research should exercise care in both how they assess quality of evidence and how they use claims about quality". However, the current systematic review does not attempt to make claims about research quality. It also exercises care in the assessment, since it does not make statements about coherence for articles which were uninterpretable for the reviewer.

7 Carroll et al. (2012) suggested that there is a case for excluding inadequately reported research articles from systematic reviews. Their study conducted a sensitivity analysis on two previous qualitative systematic reviews in which the first author was involved and found that no theme or subtheme generated by either of the syntheses depended on those studies with the most limited reporting of methodology.
not report, or inadequately report, methodological aspects of their research. A number of quality assessment instruments (or sections of them) were not suitable for the purpose of this systematic review because they would either:

- Focus on quality of research rather than quality of reporting.
- Be appropriate for qualitative research, quantitative research, or for specific study designs.
- Mix transparency with adequacy or relevance.
- Not supply a guide for users.
- Assume users of these instruments to have a considerable amount of knowledge on research methodology.
- Be lengthy and provide the possibility for evaluators to add written comments, rather than provide a quick and quantifiable means to assess\(^8\) reporting practice.

During the development of the data extraction form it was discovered that “systematic” in systematic quality evaluation is in practice often not what it implies, especially when the review includes qualitative research (Mary Dixon-Woods, et al., 2006a; Margarete Sandelowski, 2008). Judgement about reporting or research quality is systematic once the same criteria are applied to all articles. And a systematic review is systematic once it is possible to detect a system. Indexes for transparency and structure were developed to separate uninterpretable articles from those in which incoherence could be detected.

Table one shows a comparison of the approach of this systematic review with other instruments used to assess research articles. The comparison includes the prompts by Dixon-Woods et al. (2006b; 2004); an 18-point quality framework for qualitative research (Spencer, et al., 2003); the “BS checklist” (Kampen & Tamás, 2013); a reporting assessment checklist with a focus on reporting of methods used (Carroll, et al., 2012); a checklist including ten criteria for qualitative research (Kmet, et al., 2004) and a guideline which provides a summary of key issues to take into account when evaluating the quality of qualitative research (Fossey, Harvey, McDermott, & Davidson, 2002). A comparison is made on what these instruments include regarding transparency, structure and coherence. The following sections operationalise transparency, structure, and coherence together with more specific justifications for indicators included in the operationalisations.

---

\(^8\) An exception being “QualSyst” (Kmet, Lee, & Cook, 2004). A limit of “QualSyst” however is that each criteria puts together “is it present?”, “is it clear?”, “is it relevant?” and “is it justified?”.
Table 1: Positioning of operationalisation with other frameworks and checklists

<table>
<thead>
<tr>
<th>Checklist/framework</th>
<th>Transparency (1-7 detectable?)</th>
<th>Structure (1, 2, 6 detectable and links 1-6?)</th>
<th>Coherence (1-7 solidly reported?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Spencer, et al., 2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>Assumed, no appraisal question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Assumed, includes appraisal questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling process</td>
<td>Assumed, includes appraisal questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>Assumed, includes appraisal questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis process</td>
<td>Assumed, includes appraisal questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>Assumed, no appraisal question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>Assumed, includes appraisal questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Carroll, et al., 2012)</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Mary Dixon-Woods, et al., 2004) and (M. Dixon-Woods, et al., 2006b)</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Yes, no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Kmet, et al., 2004)</td>
<td>No, partial, yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>No, partial, yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>No, partial, yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>Assumed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>No, partial, yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>No, partial, yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Fossey, et al., 2002)</td>
<td>Considered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Considered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Considered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>No mention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Kampen &amp; Tamás, 2013)</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>Assumed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>Assumed, includes criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current thesis</td>
<td>Missing, unclear, clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRQ</td>
<td>Missing, present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>Missing, unclear, clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>Missing, present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Missing, unclear, clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>Missing, unclear, clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflexive account</td>
<td>Unstructured, moderate, structured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-point operationalisation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9 The framework includes the following sections: findings, design, sample, data collection, analysis, reporting, reflexivity & neutrality, ethics, and auditability. These sections include appraisal questions (AQ) together with quality indicators (QI).

10 The checklist takes study design and RQ together for assessment. 1) Study design and RQ: yes = choice of study design was given and explained. 2) DCM: yes = details of the data collection method are given, e.g., piloting, topic guides for interviews, number of items in a survey, use of open or closed items, validation, and so forth. 3) Sampling: yes = the selection of participants is described explicitly, e.g., purposive, convenience, theoretical, and so forth. 4) Analysis: yes = details of analysis method are given, e.g., transcription and form of analysis (with reference to or full description of method), validation tests, and so forth.

11 Relevant for comparisons with current thesis are the following criteria 1) Question/objective clearly described? Yes = Research question or objective is clear by the end of the research process (if not at the outset). 2) Data collection methods clearly described and systematic? Yes = The data collection procedures are systematic, and clearly described, permitting an "audit trail" such that the procedures could be replicated. 3) Sampling: Partial = The sampling strategy is not completely described, or is not fully justified. Or the sample does not include the full range of relevant, possible cases/settings (i.e., includes a convenience sample only). 4) Analysis: Yes = details of analysis method are given, e.g., transcription and form of analysis (with reference to or full description of method), validation tests, and so forth.

12 The summary distinguishes between "methodological rigor" (adequacy and transparency) and "interpretive rigor" (coherence).
2.2 Operationalisation of transparency

Transparency has been defined as the possibility to find in empirical publications “sufficient detail of the research question, design, and methods to allow an assessment” (Popay & Williams, 1998, p. 35). Carlsen and Glenton (2011, p. 1) start their paper stating that “transparency and accountability are key elements in any research report, not least in qualitative studies. Thorough reporting of methods allows readers to assess the quality and relevance of research findings.” Seale (1999, p. 468) called this auditing and notes that researchers should provide “a methodologically self-critical account of how the research was done”. Similarly, Tracy (2010, p. 841) states “transparent research is marked by disclosure of the study’s challenges and unexpected twists and turns and revelation of the ways research foci transformed over time”.

In short, several aspects of the research process should be supplied to allow consumers of research to make an informed decision on the usefulness of the knowledge presented in the article. For this systematic review, a transparent article would supply the following aspects for the following reasons:

- One or more central research questions, since useful knowledge is based on clear central research questions (Kampen & Tamás, 2013).
- One or more data collection methods, since evaluators of research should be able to make a judgement on the quality and type of data gathered.
- A description of the sampling process, since evaluators should be able to judge the appropriateness of the sampling strategies in light of the research question (Fossey, et al., 2002; Kampen & Tamás, 2013).
- Clear mention of a sample size for each data collection method, since sample sizes determine data quality and the extent to which it is possible to generalise findings.
- A description of the analysis method. If analysis is done that is not appropriate for the data gathered, the knowledge gathered is neither sound nor useful (Kampen & Tamás, 2013).
- One or more conclusions, since researchers should generate an answer their central research question in the conclusions (Kampen & Tamás, 2013).
- A reflexive account, since mention of reflectivity is important for evaluators of research to make an informed decision on the knowledge generated. For this review a reflexive account could either be (1) mention of an instrument effect on the data or results following from the analysis or (2) mention of fallacies. Regarding the latter, the ecological fallacy occurs when aggregate-level reasoning is applied to the individual level. An atomistic fallacy (Diez, 2002), a “reverse ecological fallacy” (Hofstede, 2002), occurs when a group conclusion is reached on the basis of exceptional cases. It was expected to find such discussions on the grounds that both theorists and methodologists have called for humility in ethnographic research for more than three decades, since ethnographic truths were found to be allegorical and thus inherently partial (Clifford & Marcus, 1986, p. 7).

13 A list made by Litman (2012) includes: a well-defined question; description of the context and existing information about an issue; consideration of various perspectives; presentation of evidence, with data and analysis in a format that can be replicated by others; discussion of critical assumptions, contrary findings and alternative interpretations; cautious conclusions and discussion of their implications; adequate references, including original sources, alternative perspectives, and criticism.
Table two shows the operationalisation of the seven aspects mentioned above. The operationalisation was based on (1) reporting encountered in the test phase of the data extraction form and (2) possible article level data which could be generated from individual research question, data collection and conclusion data in the analysis phase.

**Table 2: Operationalisation of transparency**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Categorization</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Clear central research question (CRQ)</td>
<td>Missing</td>
<td>Only sub questions supplied. Evolving research questions; CRQ research supplied in abstract; Implicit CRQ (research gap); At least one CRQ unclear.</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>All CRQs supplied with question mark; One CRQ supplied without a question mark without unclear CRQ.</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>All CRQs supplied with question mark; One CRQ supplied without a question mark without unclear CRQ.</td>
</tr>
<tr>
<td>(2) Clear data collection methods (DCM)</td>
<td>Missing</td>
<td>DCMs not explicitly supplied. At least one DCM supplied without application; At least one DCM is unspecified type of interview or unspecified type of observation.</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>All DCMs explicitly supplied and applied.</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>All DCMs explicitly supplied and applied.</td>
</tr>
<tr>
<td>(3) Description of sampling process</td>
<td>Missing</td>
<td>No description of selection procedure.</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>Description of selection procedures for minimum of one DCM.</td>
</tr>
<tr>
<td>(4) All sample sizes present</td>
<td>Missing</td>
<td>No sample size supplied.</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>Sample size for one DCM is missing or not explicitly supplied (e.g. use of “more than”).</td>
</tr>
<tr>
<td></td>
<td>Complete</td>
<td>Sample size for all DCMs supplied.</td>
</tr>
<tr>
<td>(5) Description of analysis method</td>
<td>Missing</td>
<td>No description of data handling after collection.</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>Description of data handling after collection (e.g. mention of transcription, mention of CAQDA, mention of grounded theory or content analysis).</td>
</tr>
<tr>
<td>(6) Clear conclusions</td>
<td>Missing</td>
<td>No CRQ; None of the conclusions link with a CRQ.</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>One conclusion incompletely answers or unclearly answers a CRQ.</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>One conclusion answers CRQ without unclear conclusions.</td>
</tr>
<tr>
<td>(7) Reflexive account present</td>
<td>Missing</td>
<td>No mention of instrument effect and no mention of fallacies.</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>Mention of instrument effect or mention of fallacies is unclear (e.g. mentioned without further discussion).</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
<td>Mention of instrument effect is clear; Mention of fallacies is clear; Mention of instrument effect and mention of fallacies is clear.</td>
</tr>
</tbody>
</table>

A transparency index was developed where the total number of clearly present aspects was multiplied by two and the total number of unclearly present aspects was multiplied by one. This sum was divided by the total score in case all seven aspects were clearly present (7 * 2 = 14):

\[
\frac{(2 \times \text{total aspects out of 7 clear}) + (1 \times \text{total aspects out of 7 unclear})}{14}
\]
2.2 Operationalisation of structuredness

Difficulty in assessment of qualitative research has been noticed three decades ago by Knafl and Howard (1984, p. 17), who pointed out that “the absence of a standard format for reporting qualitative research makes it difficult for even the methodologically sophisticated reader to assess the validity of a qualitative study”. Therefore, they proposed for different research purposes what should be supplied in the introduction, sample, measures, procedure, results and discussion part of the manuscript. While Carroll et al. (2012) mentioned transparency; Knafl and Howard (1984) mentioned the reporting format of qualitative research articles. Structuredness for this systematic review would not imply that articles needed to be formed as an experimental scientific report\(^{14}\), representing inquiry “as occurring in a linear process and findings as truths that anyone following the same procedures would also find” (M. Sandelowski & Barroso, 2002, p. 79). It should be acknowledged that there is a greater variation in style of reporting in qualitative research (Knafl & Howard, 1984).

These different styles make detecting the findings in articles more difficult than in quantitative research reports where the distinction between results and discussion is stricter (Margarete Sandelowski & Barroso, 2003). However, the ability to detect the findings, including the ability to assess the influence of the data gathered on these findings is a necessary aspect of an assessment of validity or credibility. For this systematic review, a structured article would imply transparency, but would facilitate detecting necessary aspects for assessment in the body text. The framework of Spencer et al. (2003, p. 14) was followed, in that articles should have a “structure and signposting that usefully guide the reader through the commentary”. A “useful guide” was further operationalised such that there would be an ability to support a systematic (reporting) quality assessment for a number of articles. That is, a structured article would supply an ability to detect (transparency) and separate the process (theory, method, and data) from empirical findings and implications.

As indicators of structure this systematic review assessed proportions of documents (as percentage of body text) related to research question, methods, conclusions, and implications. It also assessed the first shown locations of the research questions, data collection methods, and conclusions. Both these indicators would provide crude estimates of the importance of research questions, data collection methods, conclusions, and implications in the articles assessed. Given that the papers reviewed were primary research articles a fair amount of the articles on research methods was expected.

To make judgments on the structuredness of articles, these indicators were used together with a 3-point measurement scale. The scale ranged from unstructured towards moderately structured, towards structured. In a structured article there would be the ability to:

---

\(^{14}\) It has been argued that the experimental article structure is in fact an “ordered after-the-fact reconstructions of empirical research” (M. Sandelowski & Barroso, 2002; Margarete Sandelowski & Barroso, 2003) whereas in actuality empirical research is a rather disorderly, messy, undertaking (Bazerman, 1988; Law, 2006). This argument has been made by scholars who have argued for “a move from judgment based on epistemic criteria towards judgment according to aesthetic and rhetorical considerations” (Rolfe, 2006; M. Sandelowski & Barroso, 2002). According to these authors, form and content of reports are inseparable where “content is shaped, pruned and reworked to fit the form” (Rolfe, 2006, p. 308). This argument has in turn been criticized (Porter, 2007). Porter (2007, p. 81) points out that the inseparability of form and content “does not mean that the relationship between form and content is deterministic. No matter how stringent formal requirements, there is always room for the expression of unique content” (ibid, p. 81). He argues that “rather than undermining the possibility of research reports making factual claims, the fact that those reports are dynamic vehicles of mediation provides the very channels through which those claims can be communicated and judged” (ibid, p. 81). It is the writers’ responsibility to demonstrate that the research they are reporting has been conducted in a valid and rigorous manner, while the readers’ responsibility is to interpret the report to ascertain whether or not they are persuaded that the writer has indeed demonstrated rigor.
• Detect in the body text a research question, data collection method, and conclusion. This excluded missing data collection methods, data collection methods supplied in footnotes, and research question supplied in abstracts.
• Link the research question(s) with the data collection method(s) and conclusion(s).
• Distinguish theory from data, data from results, results from conclusions, and conclusions from implications. These distinctions would be facilitated by:
  • The article first supplying the central research question(s), the data collection methods second and conclusions third.
  • The article supplying considerable proportion of the document on applied methods (data collection, sampling and analysis).

In unstructured articles either point 1 or point 2 did not hold. In a moderately structured article point 3 did not hold. In a structured article point 1, 2 and 3 would hold.

2.3 Operationalisation of coherence

After an assessment of transparency and structure it would be possible to assess coherence, i.e. assess whether "the kind of knowledge generated in the results or presentation section does what it said it would do under the aims of the project" (Holloway & Todres, 2003, p. 347). An assessment of coherence is necessary since, during conceptualization and execution of (qualitative) research projects it could be possible that the area of curiosity, research question, data collected and data analysis grew out of alignment (Chenail, 1997). Articles could be out of alignment for example in their focus of study and the literature review; literature review and research questions; research questions and methodology; methodology and methods; methodology and findings; findings and research questions; findings and implications; or findings and limitations (Chenail, Duffy, George, & Wulff, 2011; Kampen & Tamás, 2013; Tracy, 2010). Editors of the Qualitative report have pointed out that while individual sections such as presentation of the problem, review of the literature, methodology, results, and discussion could be “constructed in a sound, logical, and structural sense” (Chenail, et al., 2011, p. 32), “the alignment of these parts into a coherent mosaic across the span of the paper” was lacking in many qualitative research manuscripts they received.

That is, in a coherent article the research questions would match the methods and conclusions. The checklist by Kampen and Tamás (2013) operated on this level of coherence, since useful knowledge is dependent on coherent (reporting of) research. In this systematic review, the assessment of coherence asked whether the relations between the distinguished parts in the assessment of structuredness were also solid. Criteria for a coherent article minimally overlapped with criteria for structuredness, for example regarding the number of conclusions and central research questions.

Based on the “BS checklist” by Kampen and Tamás (2013) this systematic review assessed transparent articles for some often occurring incoherencies. The operationalisation of coherence was such that in a coherent article:

• Conclusions would answer central research questions, where the number of conclusions which answered a central research question had to be equal to the number of central research questions.
• The ‘object of interest’ would equal the ‘object of research’ and ‘object of conclusion’ taking all research questions, data collection methods, and conclusions into account.
• The data collection methods would be appropriate for the highest reach of the conclusions, where incoherence was operationalised as:
  • A convenience sample without proportion of the document on analysis with a real world descriptive conclusion.
  • Unstructured interviews as only data collection method together with a descriptive conclusion with a reach higher than experiences.
  • Observations based on a purposive sample as data collection method with a real world descriptive conclusion little influenced by data gathered.
• The study design would be appropriate for the reach of the conclusions, where incoherence was operationalised as:
  • Case study designs with prescriptive conclusions.
  • Case study designs with predictive conclusions.
  • Cross-sectional designs with predictive conclusions.
• The study design would be appropriate for the type of research questions, where incoherence was operationalised as:
  • Single case study for a longitudinal question.
  • Case study for a causal question.
• The highest reach of the conclusions would match the highest reach of the research questions, where coherence would be in case the highest reach of the research question was higher or equal to the highest reach of the conclusion.
• The types of questions would be similar to the types of conclusions, where incoherence would be in case the number of conclusions differed from the number of central research questions or in case one type was not coherent between the conclusions and questions.
3. Data extraction form

3.1 Development

The data extraction form was based on a checklist developed by Kampen and Tamás (2013) intended for users of policy supporting research to “decide whether a piece of research begs further study or can be dismissed right away” (p. 1). Their checklist could be seen as an attempt to operationalise an “anti-bullshit agenda”\(^\text{15}\) (Silverman, 2013, p. 146) for policy supporting research reports rather than directly related to quality assessment in systematic reviews. This section presents difficulties encountered during piloting together with major adaptations after each of the four pilots.

The development of the data extraction form started with an exercise to break down the article by Kampen and Tamás (2013) to get insights into necessary variables to be extracted for identification of mismatches. Based on the result, an initial draft of variables for ‘research question’, ‘method’ and ‘conclusion’ was made. For the research question it was decided to extract data regarding the attitude towards the social movements\(^\text{16}\), the type of question (causal, correlational directional, correlational non-directional, longitudinal and descriptive) and the proportion of the document (POD) (as percentage of body text). For the conclusion it was decided to extract data regarding the attitude towards the social movements, type, POD, the source of the conclusion (empirical or theory) and the reach of the conclusion (research context, experiences, and real world). For the methods used it was decided to extract data regarding sampling method, data collection method, analysis method, and POD relating to reporting of methods used.

These variables needed to be extended with other variables for theory, results and implications in order to identify mismatches. The first pilot\(^\text{17}\) was to detect proportions of document on research question, theory, method, results, conclusion and implications. Sections of articles were double coded in order to provide information on the overlap between the sections, i.e. to detect theory in the conclusion section, research questions in the method section etc.

POD relating to research questions was initially defined as parts of the articles supplying the research question, with research questions being questions intended to generate knowledge. However, it was noted during this test that research questions were often not explicitly supplied. Indications of research questions were also given in different forms such as headings; as aims and objectives; as parts where the structure of articles were presented, or as research gaps.

Bottom up codes were added to the data extraction form in order to capture for such differences. Thus, POD relating to research questions had to be extended to being parts where

\(^{15}\) For Silverman bullshit is that which is “overly kitsch, overly jargonized, and over-theorized” (2013, p. 129). He believes that qualitative researchers habitually – and increasingly – underemphasize the empirical, resulting in work that is less theoretically alive than it might otherwise be. According to him, qualitative research has been infiltrated by current trends in contemporary culture, in particular a search for ‘raw experience’ and ‘emotionalism’. In his “anti-bullshit” agenda he calls for a renewal of traditional values such as clarity, reason, economy, beauty, and truth.

\(^{16}\) The “attitude towards social movement” was included since it could be an explanatory variable for possible mismatches identified in research articles. However, this was taken out of the data extraction form since it was highly subjective and lost significance when the review took a step back.

\(^{17}\) The test included the first six accessible articles found in an initial search in Web of Science. These articles were imported in Atlas.ti to start detecting these PODs. The first test on six articles included the codes; Body text; RQ_POD; Theory_POD; Method_POD; Results_POD; Conclusion_POD; Implications_POD. The Body text was seen as all text excluding title, abstract, footnotes, acknowledgement, introduction stories, quotes before the introduction, postscripts, headers, and references.
research questions were directly stated\textsuperscript{18}; parts explaining the article or section structure\textsuperscript{19}; research gaps\textsuperscript{20} and some headings in articles\textsuperscript{21}. The POD related to theory could give an estimation of the importance given to theory for the formulation of the conclusion and implications. However, POD related to theory could also be subdivided between clarification of concepts and variables based on previous research; positioning of the article within the research of others\textsuperscript{22} or theoretical readings of past events, i.e. different historical representations. These bottom-up codes were added for theory.

During this first pilot it was also noted that method sections were either missing or the content did not include a procedural description. Method sections included literature discussion of concepts from which a method seemingly had to be inferred, e.g. frames or discourses. A new code\textsuperscript{23} was introduced which was applied to mention of the sampling method, data collection method, and analysis method. Also, the decision was made to include footnotes as body text, since articles reported data collection methods in footnotes.

Relating to the POD of results it was noticed in the first pilot, that in order to find the results, it would first be necessary to have a clear interpretation of the research question and analysis method. In case the research questions and analysis methods were not explicitly supplied, identifying results of articles would not be reliable. It was also noted that results could be both from secondary sources and from primary sources, which made for a loose distinction between results and theory. Bottom up codes were added to capture for these two types of results.

POD relating to conclusions was initially defined as answers to the central research question. However, there were often multiple questions asked, both sub questions and central research questions. In case the distinction between sub questions and central research questions was not clear, detecting conclusions of articles was not reliable. Also, central arguments were difficult to distinguish from answers to central research questions. Therefore, a variable was

\textsuperscript{18} The code “RQ-S” was applied to stated research questions. In one test article stated questions were in the following form: “To better understand the CSR (corporate social responsibility) of ECSA (name of company), it is important to analyse the discourses, the corporate logic and the company’s actions. In other words, we need to ask the following: What is the objective of this company? What is their vision for this project? How did they propose to carry out these objectives?” (Ximena S. Warnaars, 2012 added abreviations).

\textsuperscript{19} The code “RQ-AS” was applied to research questions in the form of article structures. Christof Mauersberger (2012) supplied a research question in the form: “After retracing these two stages, a more analytical look will be given to the dynamics of media–state relations and at explanations for the success of the social movement”.

\textsuperscript{20} The code “RQ-RG” was applied for research gaps. For example: “Nonetheless, we know much less about the conditions associated with local level variation in opposition to market reforms.” (Paul Almeida, 2012).

\textsuperscript{21} In later versions of the data extraction form headings were part of article structure. At times headings presented seemingly new questions in the research articles. Some examples of headings encountered during the pilot are: “Explaining the movement’s impact: associational power and framing” (Christof Mauersberger, 2012). “Managing the tensions between state and social movements” (Benjamin Kohl and Linda Farthing, 2012).

\textsuperscript{22} For literature discussions the bottom up code “Theory_LD” was introduced. The following is an example where this code was applied. In a small section coded as being part of theory, there was a mention of a research gap and a conclusion or implication. Warnaars (2012) writes: “I situate my work within the literature on political ecology that studies resource-related conflicts with a specific geographical focus on Ecuador. Studies on struggles over natural resource have draw attention to [...] Some academics analyse and emphasise [...] whereas others bring attention to [...] Some studies of resource conflict examine [...] Less scholarly attention appears to be paid to [...] Their studies reveal that [...] The argument presented in this paper follows a similar analytical thread and adds that CSR programmes, discourses and activities should be understood as a constitutive part of social and territorial transformations”.

\textsuperscript{23} The code was named “how?”. In later data extraction forms this code was again subdivided between “how_analysis”, “what data collection method”, and “how_sampling”. An example where the code “how?” was applied in an article included in the first pilot was: “Empirical evidence has been obtained by document-based process tracing (how_analysis), complemented by 24 additional interviews conducted with movement activists, representatives of commercial media, opposition politicians and local researchers (what data collection method). Interviewees were chosen on the basis of their involvement in the debate and were contacted directly by the author (how_sampling)” (Christof Mauersberger, 2012). This quote was taken from a footnote.
introduced on the *clarity of the research question* (stated with question mark, stated without question mark, implicit, and unclear), and central arguments were also interpreted as conclusions. Relating to the POD of implications, these could also be of multiple types, such as policy recommendations, suggestions/proposals for the use of other concepts or methods, advice towards social movements, indications for future research, limitations of the research, or implications for theory. Bottom up codes were added in order to capture for different types of implications.

The next test assessed whether adding bottom up codes made coding more efficient and whether the variables in the data extraction from relating to the research questions of articles could be extracted reliably. Several problems were encountered when returning to the previous articles used in the pilot for POD. Extracting data regarding the research questions was not reliable since extracting data for the number of questions provided difficulties in articles with unclear reporting of research questions. Also, an unclear distinction between sub questions and central research questions made mechanical application of the data extraction form regarding the number of research questions difficult. Adding bottom up codes for POD made data extraction more efficient.

Development of the data extraction form continued regarding to the research methods used in articles and the influence of theory on the conclusions reached. In order to further develop the method section of the data extraction form, other sources were used such as: (a) “Unit and level of analysis” in social movement research (B. Klandermans & Staggenborg, 2002, p. xv); (b) Sampling types (Kumar, 2010, p. 198); (c) Data collection methods (Kumar, 2010, p. 139); (d) Study designs and types of questions and conclusions (De Vaus, 2001; Kumar, 2010) and (e) Steps from data gathering towards data analysis (Sapsford & Jupp, 2006).

After operationalisation of methods, it was necessary to capture the influence of the data gathered on the conclusion. In the data extraction form this was intended to be captured by: a) the proportion of document codes, b) a scaled code, c) a subset of a system proposed by

---

24 During this phase of the data extraction form development the code “Conclusion_CA” was used for central arguments, however this was not done for later data extraction forms. An example of an argument interpreted as a conclusion was: “I argue that the successful passing of the legislation can only be understood by considering the associational power of the movement behind it, its potent framing of the reform, and the changing relations between the media and the state” (Christof Mauersberger, 2012).

25 For implications for future research the code “Implications_FR” was used. The following is a quote from an article included in this first pilot: “This study maintains that analysts need to classify more precisely the local assets inside a community that are most likely to be converted into collective action in the developing world (Boudreau 1996). In order to build probabilistic models of the emergence of collective action in developing countries, it is necessary to discern which types of administrative, physical, social and organizational properties are more favourable for common people to appropriate locally and engage in social movement-type activities.” (Paul Almeida, 2012).

26 For limitations of the research the code “Implications_limitations” was used. An example of a part of an article in the first pilot was: “Although, from a democratic point of view, diversity of content is central, no satisfactory measurement of content diversity exists to date (Just, 2009). Hence figures on ownership and market concentration are used as a proxy. However, in the Latin American context, even those figures are difficult to obtain on an up-to-date basis (Becerra and Mastrini, 2009: 25).” (Christof Mauersberger, 2012)

27 Operationalisation of analysis method for the data extraction form was to a large extent done during a discussion meeting. The question “is there any mention of inconvenient data?” was used as an indication that all data gathered was included in the analysis.

28 A list of specific sampling types was used. In Atlas.ti it would be useful to start coding on the lowest possible level, the specific sampling techniques, and later merge codes towards higher/broader levels like random or non-random sampling.

29 The influence of data on the results could be (subjectively) measured on a scale: influence empirical on results %1, influence empirical on results %2, influence empirical on results %3. During this stage of the development it was noticed that Atlas.ti had the possibility to code on a (nominal) scale level which would be more efficient as it would save time during export to SPSS and prevent mixing codes given to research questions and conclusions. Using scale variables it was also possible to code the page number where there was a first mention of a research question, data
to classify theoretical claims, and d) binary questions in the theory and conclusion section of the data extraction form.

The changes made to the data extraction form were tested in a second pilot, which included one Spanish article and three English articles. However, the data extraction form did not facilitate efficient extraction of data, since the “abcdef structure” of the data extraction form was not the sequence in which it was possible to extract data from articles. First, the POD codes could not be applied to a specific section because sentences were often a mixture of theory and data gathered. Thus, extracting data regarding POD was time consuming. Second, the data extraction form did not capture for linking multiple research questions with multiple data collection methods and multiple conclusions. This made judgement on the influence of theory on the results and conclusions rather difficult. Third, there was a suspicion that research questions changed through the articles. It was noticed that some PODs relating to research question could be interpreted as descriptive while in other PODs it was correlational. Fourth, it was decided only to use a dimension of theory by Strauss called conceptual complexity, but this assumed that it was possible to identify the concepts used, together with links with other concepts. This assumption did not hold, which made extracting data regarding theory highly subjective.

Based on these efficiency and reliability problems encountered in this second pilot, it was decided to stop extracting data regarding the theory and result sections of the data extraction form. Detecting a specific theory section was not possible in most cases and detecting results could also not be done reliably since it depended on reporting of research questions, data collection methods and analysis methods.

The data extraction form had to be split into variables on an article level and variables on the level of individual research questions, methods and conclusions. Also, three main additions were made to the data extraction form. First, a question on research question origin was added to capture whether a question was completely new or whether it was a rephrasing of a question stated before. In a similar way it was decided to code the source for each individual conclusion. Second, it was decided to add a variable time needed to code article which would measure coding reliability and of the structuredness of articles. Third, a variable was added on the structuredness of articles.

The resulting version of the data extraction form was tested in a third pilot of four articles. Again difficulties were encountered during data extraction. First, in this pilot there was collection method or conclusion (FSL-RQ %pagenumber, FSL-Method %pagenumber, FSL-Conclusion %pagenumber), because the POD codes on their own were not enough to describe the structure of articles. Furthermore, scaled codes were useful in characterizing the influence of theory on conclusions reached, where everything which could not be seen as part of the empirical research was judged to be theory based.

Straus proposes six dimensions in order to crudely locate a given theory. These dimensions are: Abstraction: generality of conception; Scope: number of substantive areas studied; Range: extent of relevance Specificity: of detail of grounding; Conceptual complexity: links with other theories and concepts; Applicability: relevance to aspects of the ‘real world’. These dimensions identified by Straus were useful, but it was opted not to operationalise all dimensions into questions for the data extraction form and only extract data regarding conceptual complexity and applicability with applicability interpreted as generalizability. The dimensions of Strauss overlap and use of these dimensions assumes that it is possible to identify a single theory in an article, which was not the case.

Because it was felt that there were too many affirmative articles, the question “is this a surprising conclusion?” was added. Also the questions “Does the author make a generalization to other situations or research areas?” and “Is there a justification for the generalization?” are examples of binary questions added to the data extraction form. These questions were replaced to include mention of fallacies and mention of instrument effect.

It occurred that coding started for two articles, but a strong suspicion arose that these articles were not primary research. Other articles had to be put in place in order to test the data extraction form. During the test rounds, the four articles were fully coded for the POD codes, next coded for the other questions related to the content of specific PODs.

In order to see links between the RQ, DCM and conclusion within articles it was intended to use colours in Atlas.ti, and it was tried to make hyperlinks in the articles to show justifications of codes given.
an article in which there was no research question present in the body text. In that case it was difficult to continue coding the article. In the abstract there was an indication of a research question\textsuperscript{34} which was used for further data extraction from the article. Second, the clarity of the conclusion was noticed in this pilot. Conclusion sections could be providing a lengthy summary which made it difficult to identify specific conclusions. Third, the predefined lists of levels of analysis and unit of analysis provided difficulties. Predefined lists were not mutually exclusive and often another unit of interest had to be added to the lists. Fourth, it was noticed that there were articles in which there was no indication of who was interviewed, and some articles did not explicitly mention the type of interviews conducted or observations made\textsuperscript{35}. Fifth, the first shown location of the methods used was not specified in the data extraction form, since it could be mention of data collection method, analysis method, sampling, or study design.

The last adaptations to the data extraction form were made in order to extract data efficiently and reliably. First, it was decided to number each question, data collection method, and conclusion and to extract data regarding the page number where these three aspects were identified respectively\textsuperscript{36}. Extracting data regarding the page number of identification could provide a measurement of the spread of these aspects over articles. Second, numbering the individual questions, data collection methods, and conclusions, facilitated making a distinction between sub questions and central research questions. A variable was added for each individual conclusion whether it answered a sub question or the central research question. Third, it was decided to adapt the level and unit of analysis variables in the data extraction form towards a similar list of “object of interest”, “object of research” and “object of conclusion”; which had to be applied to each research question, data collection method, and conclusion respectively.

After applying these last changes, the first ten articles were sampled and the data extraction form was applied to all ten articles. After coding the first ten articles, some minor changes were made to the data extraction form and the first ten articles were re-coded\textsuperscript{37}. First, because all ten articles were case studies, a subdivision was made between comparative, longitudinal, and multiple case study designs. Second, articles which did not mention a sampling strategy were judged to be using a convenience sample. Third, some articles had partial answers to central research questions. Last, it was noticed that articles mentioned a data collection method without application in the article. It was therefore decided to add the option ‘over’ for each data collection methods, the link between conclusion and research question, and for the variables ‘mention of fallacies’ and ‘mention of instrument effect’. The option ‘over’ was defined as overrepresentation of the empirical work reported in the article.

\textsuperscript{34} The abstract supplied a question in the form: “This case examines the municipal housing policy from its early days until the present, involving considerable articulation with social movements, and the role played by different community activists from the church and the trades unions” (Kazuo Nakano et al., 2009). From this sentence in the abstract it was possible to code the “level of analysis” of the research question (later to be called “object of interest”); the level of clarity; the reach of the question, and finally it could be used to code for the study design as in the article there was no explicit mention of it.

\textsuperscript{35} Two examples of unclear reporting of the type of interviews were: “Settlers’ experiences of the countryside as it emerged from the interviews [...]”(Rute Caldeira, 2008); “Part of the research for this case study was undertaken by interviewing housing movement leaders and government actors involved in the urbanisation processes and struggles over the last 20–25 years.”(Kazuo Nakano, 2009).

\textsuperscript{36} It was noticed in a later phase that numbering the individual research questions, data collection methods and conclusions was a useful decision to make. Each variable for which data had to be extracted on an individual question, data collection method or conclusion level had to be linked to the number of question, data collection method or conclusion respectively. This was necessary for export from Atlas.ti towards SPSS. However the data regarding the page number of identification of each individual RQ, DCM and conclusion was lost in this process.

\textsuperscript{37} Further changes on the content of the data extraction form were not made during the data extraction process, with the exception of code names to be in the form of question %answer. This was necessary for an export of data from Atlas.ti towards SPSS.
3.2 Application

Data was extracted in the following order for all ten articles in the cohort:\(^{38}\); 1) body text and page numbers; 2) research questions; 3) data collection methods; 4) conclusions and implications and; 5) the study design and the degree of structuredness. First the body text and total number of pages with body text were coded for all ten articles in the cohort. Body text\(^{39}\) was defined as total word count of the article excluding the title, abstract, acknowledgements, introductory stories, quotes above the introduction, and references. Footnotes were inspected individually to be included in the body text. Tables, figure headings, and figure notes also belonged to the body text of articles. After body text was coded for, the total number of pages from introduction until the references was counted. Half pages with body text were counted as complete pages.

After extracting data regarding body text and the total number of pages with body text, data regarding the research questions were extracted for all articles in the cohort. First, the proportion of the document (POD) on research questions was detected. PODs could either be stated questions, research gaps, or parts describing the structure of articles. Second, the page number of the first location of a research question was extracted.\(^{40}\) The first shown location of a research question was defined as the first mention of a stated question. In case a stated question was missing in an article, a research gap was identified for the first shown location of a research question. The abstract was used to extract data regarding the research question in case neither a stated question nor a research gap could be detected.

Third, the POD sections were examined in order to number each individual new question or inadequate rephrasing of the central research question. This third step also implied extracting data regarding the origin of research question (central research question, sub question, new question, adequate rephrasing or inadequate rephrasing). Adequate rephrasings were not numbered and no further data extraction occurred for adequate rephrasings. For articles which formulated the research question as a research gap rather than a stated question, the research gap was identified as central research question and numbered. For articles with both a stated question and a research gap, there was – except for POD – no further data extraction regarding the research gap.

After numbering the research questions together with the research question origin, data was extracted for each research question individually regarding:

- The type of research question. Types could be causal, correlational (directional, non-directional), descriptive or longitudinal.
- The clarity of the research questions. Clarity could be implicit (for research gaps), stated with a question mark, stated without a question mark or unclear.

\(^{38}\) The final data extraction for was inserted in Atlas.ti such that the questions were families in the code manager and the answers were individual codes. The code manager was ordered in sections from research questions towards conclusions and implications. It was possible to use filters, which were used in order not to be influenced by data extracted in previous rounds of coding. Also, the aim was to code 20 articles per week. This high number was taken in order to apply codes mechanically rather than take too much time coding an individual article.

\(^{39}\) Because articles often had headers including the name of the journal and page number, the code could not be applied once per article but had to be applied per page of the article.

\(^{40}\) The page number was not the page number in the journal, but it was counted from the first part of body text downwards in the similar way as the total number of pages was counted.

\(^{41}\) In case the abstract was necessary to extract data regarding the research question the code “FSL %99onlyinabstract” was used.
• The reach of the research question. Reach could be the research context, experiences\textsuperscript{42}, imaginations of the real world\textsuperscript{43}, interpretations of the real world\textsuperscript{44}, how the real world looked like in the past\textsuperscript{45}, a description of the real world, on the real world in the future\textsuperscript{46} or real world prescriptive.

• The object of interest. Object of interest was divided between individual object of interest and aggregate object of interest. Multiple codes could be attached to the same quotation to distinguish for example between social movements within a municipality and social movements within a nation. If the object of interest was not in a predefined list (see appendix 1), the object of interest was then coded as other and the specific object would be inserted manually. The same process was applied for individual object of interest as for aggregate object of interest.

After extracting data on the research questions, data was extracted regarding the research methods of the ten articles in the cohort. First, all areas in the article which belonged to the technical aspects of the methods used were coded as the proportion of document relating to methods. That is sections which supplied what data collection methods were used, what sampling procedures were used and how analysis of data proceeded. Second, the first shown location of a data collection method\textsuperscript{47} was detected. The first shown location was defined as the first mention of interviews, observations, document review etc. In case there was no explicit mention it was applied to the page where it was possible to detect the use of a data collection method. The data collection methods were numbered\textsuperscript{48} after which it was checked whether all data collection methods had a sample size reported. Sample sizes could be missing for all data collection methods, missing or unclearly reported for one data collection method, or reported for all data collection methods. Also, it was detected whether articles supplied a reflexive account. A reflexive account could be mention and discussion of inappropriate generalisations\textsuperscript{49} (mention of fallacies) in the field or in the related research. Inappropriate generalisations could be from data gathered at individual level towards making collective inferences, or from data gathered at a collective level towards making individual inferences. Articles in which an uninterpretable mention of fallacies was detected, or which did not further discuss the

\textsuperscript{42} An example of a question with the reach experiences: “This article examines women’s ways to poner el cuerpo in political resistance and the meanings attached to these actions” (Sutton, 2007).

\textsuperscript{43} An example of an imagination of the real world: “how residents of popular municipalities imagined democracy” (Greaves, 2004).

\textsuperscript{44} An example of a real world interpretation reach was: “In analyzing this movement I shall be concerned with the variety of interpretations and reinterpretations that the actors made of the events in which they participated” (Nash, 1992). Experiences, imaginations and interpretations were collapsed in the analysis phase to be part of experiences.

\textsuperscript{45} An example of a question with a real world in the past reach was: “Drawing on primary and secondary sources, this article discusses the durability of communist ideology in rural Brazil during the second half of the twentieth century. It analyzes the theme in two major periods: the Populist Republic (1945–1964) and the Military Regime (1964–1985)” (Welch, 2006)

\textsuperscript{46} Two examples of predictive questions were: “But does all this mean that there are few prospects for true democratic reform in Mexico’s future?” (Davis, 1994) and “In the final part, the future of rural social movements are analysed and relevant conclusions are drawn” (K. B. Ghimire, 2003)

\textsuperscript{47} The code “FSL-DCM %” was used where the page number of body text was entered after the percentage sign. The code “FSL-DCM %99only in the footnotes” was applied to articles which supplied the data collection methods in the footnotes. The code “FSL-DCM %98none” was applied to articles which did not explicitly supply a data collection method.

\textsuperscript{48} Semi-structured interviews and focus groups were also categorized as unstructured interviews. Therefore, it would be possible to have more than one data collection methods coded as unstructured interview.

\textsuperscript{49} Inappropriate generalisations could be from interviews with leaders towards that of the movement, or from data collected on the movement level towards individual activist level. The name of the fallacy (ecological fallacy, exception fallacy, atomistic fallacy, reverse ecological fallacy) did not have to be mentioned.
implications of the inappropriate generalisation were coded as having an unclear mention of fallacies. Another form of a reflexive account could be the mention of an instrument effect. An instrument effect could be an effect of the data collection method on the data gathered or an effect of the analysis method on the results.

After numbering each data collection method, data was extracted for each individual data collection method regarding:

- The specific data collection method, which could be unstructured or structured interviews, participant or non-participant observation or a document review. Articles which reported a data collection method as “interviews” or “observations” without supplying that they were unstructured, structured, participant, or non-participant, were coded as unknown kind of interview or unknown kind of observations. Articles for which data collection methods were detectable without them being explicitly stated or vice versa were coded as overstatements.\(^50\)
- The sampling type, which could be random, convenience, or purposive sampling. The code convenience sample was also applied to cases where a sampling strategy was not supplied.
- The object of research\(^51\), separated between individual object of research and aggregate object of research. The same process was applied as for extracting data regarding the object of interest. Multiple codes were applied to the same quotation since unstructured interviews were conducted with individuals from different backgrounds. It was necessary to separate individuals within a movement from those individuals outside the movement. In case the supplied object of interest was not included in the data extraction form, the object of research would be coded as other and the specific object would be inserted manually.

After extracting data regarding the data collection methods, data was extracted regarding the conclusions and implications\(^52\) of all ten articles in the cohort. First, the conclusions reached in articles were detected and coded as the POD related to conclusions. This step also implied making a distinction between implications and conclusions. PODs related to implications were detected and coded for. Implications could be: policy recommendations\(^53\); social movement recommendations\(^54\); (future) use of concepts\(^55\); future research\(^56\); limitations of the reported

\(^50\) Unknown kind of interviews and unknown kind of data collection methods were collapsed with overstatements to be unclear data collection methods.

\(^51\) Objects of research compared to the objects of interest could be spread across the article. They were coded together with the number of data collection method.

\(^52\) Since it was difficult to distinguish conclusions from implications it was decided to extract data regarding both in one round of coding.

\(^53\) A policy recommendation could be found in Grabe (2012): “In other words, the design of empowerment programs should be based on the potential for transformative change and on outcomes that suggest a greater ability on the part of women to act on the structures of power that constrain their lives, and not on buzzwords that are in favor of a neo-liberal globalized economy. These designs may be most effective when policy makers and interventionists work with women’s organizations to combine equity in the distribution of resources with a sense of personal power and control to optimally impact wellbeing.”

\(^54\) A social movement recommendation could be found in McCormick (2006): “Finally, this research offers new ideas for building successful environmental movements in other contexts. Environmental movements are the classic case to contest expert knowledge because related policy is clearly based on science. Many other activists around the world are using tactics similar to anti-dam movement activists. Understanding these innovative methods may contribute to the development of more effective environmental movements in many settings.”

\(^55\) An implication for the use of concepts could be found in Bebbington (2010): “This does not mean that movements are irrelevant to debates on poverty. However, tracing the linkages between the two domains requires a conception of
research, and theoretical implications.

Individual conclusions were numbered and the first shown location of a conclusion was detected. For each individual conclusion, data was extracted regarding:

- The **link the conclusion had with the research question**, which could be an answer to a CRQ, an answer a sub question, no link at all with a research question, a partial answer, or an unclear answer to a research question.
- The **source of the conclusion**, which could be empirical, theoretical, both empirical and theoretical, or not specified at all.
- The **type of the conclusion**, which could be causal, correlational (directional, non-directional), longitudinal or descriptive.
- The **reach of the conclusion**, which could be the research context, experiences, real world imaginations, real world interpretations, real world in the past, real world description, predictive, prescriptive.
- The **object of conclusion**, divided between individual object of conclusion and aggregate object of conclusion. Data extraction regarding the object of conclusion did not differ from data extraction regarding object of interest and object of research.

Last, the **influence of data gathered on the conclusion** was coded on a three point scale (high, both theory and empirical, low). This was coded on an article level rather than on the specific conclusion level. After extracting data regarding the conclusions, the last round was to extract data regarding two ID variables for all ten articles in the cohort. These two variables were **structuredness**$^{57}$ and the **study design**. The degree of structuredness could be low, weak, moderate or high. The study design could be a case study (single$^{58}$, comparative$^{59}$, longitudinal$^{60}$, multiple$^{61}$), cross-sectional$^{62}$ or longitudinal.

### 3.3 Efficiency, reliability and internal validity

A number of decisions made data extraction more efficient compared to the different pilots. The first was the decision to stop extracting data regarding the theory and results of articles. Second, limiting proportion of document on methods used to the technical aspects of the methods used

poverty as more than income based, of the causes of poverty as rooted in relationships of power, and of poverty reduction policy as determined by political and discursive processes in which movements are embedded (see Mosse, 2010).”

$^{56}$ An implication for future research could be found in (Grabe, 2012): “Nevertheless, future research should aim to construct measures that will more accurately capture women’s individual agency”

$^{57}$ During data extraction a 4-point scale was applied. This was later collapsed to a 3-point scale, taking low and weak together as unstructured.

$^{58}$ A single case study was conducted by Adams (2000): “Because I carried out the fieldwork in 1995-1996, readers might want to categorize my methodology as oral history informed by participant observation”. The tile of the article was Movement socialization in art workshops: A case from Pinochet’s Chile.

$^{59}$ A comparative case study was conducted by Vom Hau (2010); “The fourth section moves to the provincial level and compares Diaguita and Mbya mobilization around land and territory, followed by a discussion of the implications of these mobilising efforts for citizenship and poverty”.

$^{60}$ An example of a longitudinal case study design is in an article by Babb (1997); “My research in Managua has focused largely on the first of these developments, as I have followed the changing fortunes of a number of women (and some men) who work in small industries and commerce in the city”.

$^{61}$ An example of a multiple case study design can be found in the title of an article by Wolff (2009); “De-idealizing the democratic civil peace: on the political economy of democratic stabilization and pacification in Argentina and Ecuador”.

$^{62}$ Schmitt (2010) uses a cross sectional design; “When analysing the sources of civic engagement, several factors have to be taken into account. The following section discusses the influence of socioeconomic, socio demographic, political and cultural determinants on social participation in Ecuadorian rural communities.”
made coding the questions related to content (type, reach, clarity, origin) also more efficient. Third, the strategy to go through articles in phases made time per articles considerably less, since this strategy limited reinterpretation of questions and conclusions. Less efficient was coding the “objects of interest”, “object of research” and “object of conclusion” since most articles had multiple questions and multiple data collection methods spread throughout articles.

Reliable use of the data extraction form was dependent both on reporting practices of the articles included in this review and on the interpretation of text by the reviewer. Unclear reporting of (methodological) aspects of the research was kept track of. While further data extraction was conducted, care had to be exercised in making claims about the coherence of these articles. The strategy chosen to code in different phases was intended for reliable data extraction. Questions with a high degree of subjectivity were limited during the development of the data extraction form. For those highly subjective questions which remained, a users’ guide was included in the data extraction form used.

Non-reporting or unclear reporting of methodological aspects in articles also influenced the internal validity of the data extraction form. What was being measured in fact was the authors’ inability to interpret articles rather than what the data extraction form intended to measure which were important characteristics of the research question, data collection method and conclusion for an assessment of coherence. The decision to take a step back towards transparency and structure was made to re-align measurement to what was possible to measure.
4. Sample

4.1 Search syntax

The research population was defined as *English and Spanish social science primary research articles on social movements in Latin America*. No other criteria than the abstract, keywords, and title were used for an article to classify as social movement research. An article could also be included to the population in case ‘urban movement’, ‘peasant movement’ or ‘student movement’ appeared in the keywords, abstract or title. Both English and Spanish articles were included to generate a realistic population of Latin American social movement articles. Reading Spanish articles would not provide difficulties.

The search involved a Boolean search in the electronic database Web Of Science (Boolean search explained in Branley, 2012). The syntax was developed in Web of Science and Scopus. It was possible to detect which keywords were influential and which strategies limited non-population articles by conducting a variety of searches, scanning the abstracts and titles of articles which were not in a previous search.

Table one presents the final search syntax as it was inserted, with no time limit set, on January 16th 2013 in ISI Web of Science. The search resulted in 549 references for 510 English and 39 Spanish articles.

63 The initial strategy was to conduct a search in both Web of Science and Scopus. These databases complement each other and neither database is inclusive (Burnham, 2006). Initial searches included the search terms “Latin America” and “Social Movement”. The search expanded to include all individual countries to which Latin America referred. In order to limit the number of non-population articles other terms expected to be used by scholars in the field of social movement research were included to the syntax (e.g. ‘favela’, ‘barrio’, ‘squatters’, ‘peasants’, ‘urban’). Initial searches also included specific movements, but the decision was made not to include names of movements in order not to end up with a sample on one specific movement. Furthermore, in order to limit the search to social science articles attempts were made to include research domains and research areas, but dropped again later. The Boolean expression Near was included in the syntax since scholars use other names for social movements. Since the aim of this review was exploratory it was not necessary to use both databases. For several reasons it was decided to limit the search to Web Of Science:

- While Scopus covers a wider journal range, it is limited to recent articles compared with Web of Science (see also Falagas, Pitsouni, Malietzis, & Pappas, 2008). Web of science could provide a greater number of articles dating back several decades.
- ISI Web of Science offers the possibility to search in the Social Science Index. To specifically search for social science articles in Scopus is more complicated.
- It was observed that the same reference could have different ‘citations per year’ when comparing Scopus and Web of Science on this issue. Also, information of citations per year in Scopus would be available only for articles published after 1996. Since citations per year could provide interesting data, it was practical to limit the search to Web of Science.
- The syntax was developed only to include primary research articles. However, a reference called “original article” in one database could be a “review article” in the other. Also, a “conference proceeding” in one database could be an “article” in the other (see also Li, Burnham, Lemley, & Britton, 2010). Based on this observation it could be expected that a higher number of articles would have to be scanned for abstracts which would have been time consuming.
- In order to avoid the time consuming task of not having to deal with duplicate references it was practical to use one database.

64 On January 23rd extra individual searches for Spanish articles were done in order to see if Spanish articles were systematically excluded. Terms were for example ‘América Central’ and ‘movilización’. These extra searches did not generate more articles to be included in the population.
Table 1: Search syntax

<table>
<thead>
<tr>
<th>Population requirement</th>
<th>Search terms used*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Social movements</td>
<td>TS=(&quot;s*ocial NEAR/3 *ovement&quot; OR &quot;ovimiento&quot;)</td>
</tr>
<tr>
<td>(2) Latin American**</td>
<td>AND TS=(Mexico or Guatemala or Honduras or Nicaragua or &quot;El Salvador&quot; or Belize or &quot;Costa Rica&quot; or Panama or Colombia or Venezuela or Ecuador or Surinam or Peru or Brazil or Bolivia or Paraguay or Uruguay or Chile or Argentina or &quot;Latin* Americ*&quot; or &quot;South Americ*&quot; or &quot;Central Americ*&quot;)</td>
</tr>
<tr>
<td>(3) Social science</td>
<td>Databases=SCI-EXPANDED, SSCI, A&amp;HCI***.</td>
</tr>
<tr>
<td>(4) English and Spanish</td>
<td>AND Language=(English or Spanish)</td>
</tr>
<tr>
<td>(5) Primary research</td>
<td>AND Document Type=(Article)</td>
</tr>
</tbody>
</table>

*The asterix is a wildcard. For example "Latin* Americ*" can be Latin American, Latin America and Latino America. The asterix (*) was used to include capital letters and social movements (in the plural). The Boolean expression OR makes a combined set containing at least one of the search terms, and the expression AND makes a set consisting of elements that contain both subsets.

**Latin America was defined as the countries in this list, meaning from Mexico downwards excluding French and British Guyana and Caribbean islands. Articles were also included in the population if they analyzed more than one of these countries or made a comparison of a movement in a Latin American country with one in another continent.

***SCI-EXPANDED= Science Citation Index Expanded (1945-present); SSCI= Social Sciences Citation Index (1956-present); A&HCI= Arts & Humanities Citation Index (1975-present)

4.2 Classification

The references of the search result (N=549) were downloaded to Endnote. After deleting duplicates, a scan was conducted for the inclusion criteria of table 1 based on the titles, keywords, and abstracts. Articles were classified to the groups “population”, “population, no access”, “not social-movement”, “secondary research” and “not in Latin America”. This section provides some remarks on the classification phase of this systematic review, since there was no separate documentation for articles rejected.

References without an abstract were often classified as population since there would not be enough information for an informed classification decision65. The abstract and keywords were most influential in the final classification decision. Articles were often classified as non social movement related research when ‘social’ was near ‘improvement’ or when the topic was migration. Also, articles were classified as non social movement research when the specific social movement played a minor role in the abstract. This could be the case when a reference to social movements was made to argue that a topic was often debated or when participants in a movement were sampled for non social movement research related purposes. Furthermore, articles on the political party MAS (Movement towards Socialism) in Bolivia were not seen as social movement related, unless social movements was included in the keywords. Political parties were not seen as social movements, unless social movement was included in the

65 Itzigsohn and Von Haum (2006) did not provide an abstract. While the title Unfinished imagined communities: States, social movements, and nationalism in Latin America indicates secondary research, this was sufficient evidence to reject the article on the grounds of it being secondary research.
keywords. The often made distinction in literature between a social movement and an organisation was not explicitly made during classification.

The primary research requirement was often identified in mention of "analysis", "fieldwork", "primary sources", a specific data collection- or analysis method, or mention of a specific "case". A specific case was however not a population requirement. Indicators for secondary research were references to historical events or broad and theoretical titles. Mention of "essay" also indicated secondary research.

After classifying all references in Endnote, a list of 254 relevant articles for the review remained. Endnotes' search option was used to retrieve articles in this group. Articles which could not be retrieved using Endnote were retrieved through individual searches on January 23rd in Web of Science using access of the Wageningen University. There was no access to 34 articles and one article did not adhere to the language restrictions. A total of 219 articles remained as population for this systematic review.

4.3 Systematic random sampling

The 219 retrieved articles were imported as Primary Documents (PDs) in Atlas.ti. The PDs got primary document numbers based on the name of the documents. Articles retrieved by Endnote had document names based on their reference. Articles retrieved from Web of Science had document names in the form "name, year", based on the full name rather than surname.

Articles were systematically random sampled in cohorts of ten. Each 21st article would be sampled. A predefined list with PD numbers for each cohort was missing. For each round of sampling it was checked which article was last selected. However, the systematic random sampling strategy was not solidly executed, since counting errors were made. Sampling was random nevertheless, since articles were not looked into beforehand and articles were ordered based on alphabet rather than specific characteristics. Errors and reasons for them were recognized in the write-up phase of this systematic review. Most important errors were two counting errors. The first counting error was a minor error which did not influence further sampling. The second counting error was an error where instead of taking an interval of 21 articles one of 11 was taken. From this position the other articles were sampled again in intervals of 21. Thus, the second error influenced further sampling.

For an indication of the reliability of the classification decisions, another sample was taken from those articles classified as secondary research. An unconventional sampling

---

66 Examples of broad and theoretical titles indicating secondary research were: *Latin America's Left Turns: an introduction* (Beasley-Murray et al 2009); *Social movements, hegemony, and new forms of resistance* (Vanden, 2007); *Meanings and mobilizations: A cultural politics approach to social movements and states* (Rubin, 2004); *Agrarian Reform and Social Movements in the Age of Globalization: Latin America at the Dawn of the Twenty-first Century* (Teubal, 2009).

67 Articles downloaded from Web of Science had numbers as document names. The document names were edited in order to correctly attach files to Endnote.

68 This strategy would facilitate to stop coding articles after time available for coding passed. The aim was to sample 20 articles a week in order to extract data reliably. However, due to intensive data extraction, this aim was not reached.

69 Primary document numbers changed when one article changed position during coding. Article PD number 197 was a Word document and had to be changed towards a PDF-document. This article changed position towards number 219 and the articles with PD numbers higher than 197 get one PD number lower.
strategy\textsuperscript{70} was used, sampling articles based on the Endnote search option. This resulted in eight articles\textsuperscript{71}. A second sample of 10 articles was systematically drawn from the 93 articles, sampling each 9\textsuperscript{th} reference (see appendix 3).

\textsuperscript{70} It was first thought necessary to retrieve all 93 articles in this group in order to take a sample – in the same way as done for the 219 population articles; first exporting to Atlas.ti and then taking a sample based on their PD numbers. However such a strategy was not necessary.

\textsuperscript{71} Altieri, 2011; de Oliveira, 2010; Edelman, 2009; Hecht, 2012; Martinez-Torres, 2010; McMichael, 2008; Mitlin 2008; Puig, 2010.
5. Results

5.1 The sample

From 70 primary research articles sampled six were excluded (Figure one).

![Diagram showing the process of article selection and exclusion](image)

**Figure 1: Papers included and excluded.**

Articles in the population (N = 219) appeared in a total of 116 journals and articles to which the complete data extraction form was applied (N = 64) appeared in a total of 40 journals. Figure two shows how the latter group of articles was distributed over the journals in which they were published. The journal with most published articles in the population was *Latin American Perspectives* (21 out of 219 publications). Articles to which the complete data extraction form was applied were also mainly published in this journal (8 out of 64).

Articles in the population had a mean of 0.57 citations per year (with standard deviation 0.85). Articles to which the complete data extraction form was applied had a mean of 0.54 citations per year (with standard deviation 0.60). From the articles to which the complete data extraction form was applied, almost three quarters of the articles (47 out of 64) applied a case study design. Almost a quarter of the articles applied a cross-sectional design (15 out of 64), and two articles applied a longitudinal design.
In the final set of 64 articles, the first publication dated back to 1991 (Figure three). There has been an increase in number of publications of research on social movements in Latin America, with most articles published between 2007 and 2010.
Figure 3: Frequency counts of years of publication for the articles from search-syntax, population, and full coding scheme applied.
5.2 Transparency

Almost three quarters of the articles (47 out of 64) had multiple research questions from which 12 articles had five or more research questions. In 43 articles there was a single central research question supplied, 12 articles supplied two central research questions and 6 articles supplied three central research questions. In 59 articles a central research question was supplied before a sub question. After the first identification of a central research question however, inadequate rephrasings or new central research questions appeared in the articles.

Overall, one-third of the articles (23 out of 64) were not transparent regarding the research questions. This includes three articles which did not supply a central research question and 20 articles with unclear research questions. The three articles which missed a central research question supplied sub questions without an overarching question, such as Taylor (2011, p. 422): “[...] this article opens with a summary of the historical antecedents to the San Marcos–Condebamba Valley mobilization, before proceeding to analyze the movement’s strategy and tactics, internal organization, the problems activists encountered and moves made to surmount these. The discussion also engages with relevant literature on the characteristics of other anti-mining protests in Peru (particularly Majaz), as well as considering their wider implications vis-à-vis governance, political legitimacy and shifting power relations within the Peruvian state. It concludes with an assessment of how the latter might affect the chances of successful collective action by rural people”. Unclear central research questions supplied for example a description of a theme rather in the form; “This paper examines participatory governance in Chile through the prism of community development meetings between municipal officials and grassroots community leaders in Huechuraba [...]” (Greaves, 2004, p. 204). Unclear questions could also mix a question with a (theoretical) argument in the form; “Two main goals are pursuit. First, it will be shown how activists in the two networks mobilised emotions strategically and performed what Hochschild (1983) and Taylor & Rupp (2002) term ‘emotional labour’, or the acts of ‘channelling, transforming, legitimating and managing one’s and others’ emotions and expressions of emotions in order to cultivate and nurture the social networks that are the building blocks of social movements’ (Taylor & Rupp 2002, p. 141)” (Bosco, 2007, p. 546).

The remaining two-thirds (41 articles) supplied clear central research questions, divided between nine articles which supplied all central research questions with a question mark and 32 articles which supplied one of the central research questions without a question mark. Clear central research questions with a question mark were for example detected in Kaup (2008, p. 1734); “I ask: (1) how have transnational energy firms, the Bolivian state, and the country’s social movements worked to materially manipulate and discursively construct the country’s natural gas in order to negotiate the regulatory frameworks surrounding it; and (2) what have been the results of these actions?”. Another example of a clear central research question with a question mark was detected in Khasnabish (2004, p. 256-257); “The question of how to constitute a ‘movement of movements’ might be considered as follows: how might peoples separated by geography, distinguished by multiple political, social, cultural and economic dimensions, and who espouse a wide range of agendas, tactics and ultimate goals find the common ground necessary to generate a common front of opposition and alternatives? This question serves as the inspiring force behind this article”. Clear central research questions without a question mark were detected in Nash (1992, p. 275); “In analyzing this movement I shall be concerned with the variety of interpretations and reinterpretations that the actors made of the events in which they participated”. Clear central

---

72 Including central research questions, sub questions, new questions, or inadequate rephrasings.
research questions without a question mark were also detected in Delgado (2008, p. 561): “This paper explores how old and new representations of expert and lay knowledge shaped participation within the movement. It describes how expertise was recognized and redistributed, attending particularly to the role of MST technical coordinators and technicians”.

Regarding the transparency of data collection methods used, one article missed data collection methods. From 63 articles in which data collection methods could be detected, 45 applied multiple data collection methods. The maximum amount of data collection methods applied was four, which was the case in five articles.

Overall, two-thirds of the articles (43 out of 64) were not transparent regarding the data collection methods. This includes 12 articles73 which did not explicitly supply a data collection method and 31 articles which supplied unclear data collection methods. Unclear data collection methods did not mention whether interviews were structured or unstructured. Ghimire (2003, p. 42) for example described the data collection methods the following way; “During 1998-2000, the author had the possibility of undertaking brief field visits and participating in observant surveys in Brazil, Egypt and Nepal. Young farmers, youth groups and informed authorities were interviewed”. It could also be unclear whether observations were non-participatory or participatory. McCormick (2011, p. 38) for example described the data collection method the following way; “Ethnographic observations were conducted of movement events and meetings, government forums, and discussions between advising experts and movement groups”. An example of an article with clear data collection methods was Stephen (1995, p. 815) describing the data collection methods as follows; “[...] seven members of CO-MADRES were interviewed in depth, and meetings were observed. Maria Teresa Tula was interviewed periodically over a two-year period from 1991-93 for a total of 40 hours. Some of the material from those interviews is published as her testimonial (see Stephen 1994). Other sources of information included documents from the CO-MADRES offices in San Salvador and Washington, DC, and interviews I conducted in 1991 with women activists from seven other organizations with whom I discussed the CO-MA-DRES”.

In 83% of the articles (53 out of 64) there was no attention at all paid to the sampling strategy used (random, non-random). Examples of articles which supplied such descriptions of sampling were:

- “[...] a combination of snowballing and convenience sampling was used [...]” (Sutton, 2007, p. 130)
- “Utilizing a theoretically informed network sampling strategy, I interviewed [...]” (Kaup, 2008, p. 1735)
- “[...] this research was conducted in collaboration with a women’s organization legally facilitating women’s land ownership to obtain a sufficient number of land-owning women for the first group. The second group of women was selected from neighbouring communities in the same municipality [...]” (Grabe, 2012, p. 238)
- “In each settlement, we sampled households. Because these settlements had been regularised by INCRA, they had lists of residents, so we sought to sample based on equal probability. Random sampling, however [...]” (Perz, 2010, p. 466)
- “The communities in Cotopaxi were selected for the following reasons. [...]” (Schmitt, 2010, p. 1443).

---

73 These 12 articles include 1 without a data collection method and 11 articles for which a data collection method was identified in references or inferred from a description in the article.
Half of the articles (31 out of 64) did not report any sample size at all, and almost half of the articles (30 out of 64) had incompletely reported sample sizes. In total three articles clearly supplied a sample size for each data collection method applied. Examples of incompletely reported sample sizes were:

- “For this article, I draw on more than 45 “ethnographic interviews” conducted with “key informants” [...]” (Stewart, 2008, p. 238)
- “I conducted 105 structured and open-ended interviews with [...]” (Dosh, 2006, p. 50 in the footnotes)
- “[...] I collected data on [...] through interviews, participant observation, archival research, and analysis of published documents (e.g., scientific papers, newspaper articles, and activist websites). Over 50 semi-structured, in-depth interviews were conducted with [...]” (Kinchy, 2010, p. 507).

Almost 60 percent of the articles (36 out of 64) dedicated no attention at all to the analysis method used. A total of 8 articles supplied both a description of analysis and sampling. Descriptions of analysis methods were:

- “[...] I analyzed these data using a grounded theory approach (Glaser and Strauss 1967), coding both the participant observation notes and interviews myself.” (Adams, 2000, p. 622)
- “[...] a content analysis of the transcribed interviews was conducted” (Urkidi, 2011, p. 558)
- “the approach of this research has been to take primary qualitative data from the victims and seek to match it against corporate and/or official state accounts” (Higginbottom, 2008, p. 165)
- “Through the use of constant comparative analysis, I developed a set of working frames (Strauss & Corbin, 1994; Gamson, 1992) [...]” (Hopke, 2012, p. 371)
- “[...] transcriptions were analyzed using the QSR NVivo software package. Initial themes were drawn from the interview questions and further themes and codes were developed over the course of analyzing the data. Through this program, I was able to discover differences and similarities across interviewee type. [...] Notes were taken on each observation and also entered into NVivo to be analyzed with the interviews.” (McCormick, 2006, p. 329)
- “The content of 14 interviews was used as background information. The analysis focuses on a series of 36 interviews [...] The interviews and group of discussion focused upon two general topics [...] The information referring to farmers doing conventional agriculture was collected in a number of informal conversations with these farmers. The ethnographic data as well as the data of the interviews and discussion were reduced and analyzed following four general categories (expert knowledge, lay knowledge, ecological, non-ecological) and later organized under four subcategories” (Delgado, 2008, p. 562).

Two-thirds of the articles (43 out of 64) did not supply a reflexive account. This includes 30 articles which neither mentioned fallacies nor mentioned instrument effects, and 13 articles which supplied either an unclear mention of fallacies or mention of instrument effect. In 60

---

74 Informal conversations were not numbered as data collection method.
percent of the articles (38 out of 64) there was no mention of instrument effects on the data or results. A quarter of the articles (15 out of 64) clearly supplied an instrument effect on the data or results. For almost 20 percent (11 articles out of 64) an unclear mention of instrument effects was detected. For example, an unclear mention of instrument effects was detected in Bosco (2007, p. 548); “The same advantages that facilitated getting to know activists presented me with some unexpected challenges. In particular, the Madres questioned my outsider status as someone who had left the country, as someone who is somewhat disconnected from daily life in Argentina, and as someone who writes about social movements but who is not much of an activist himself. I was presented with difficult questions that made me reflect more critically about my positionality, my politics, and the motivations for the research (Rose 1997; Pratt 2000).” An example of a clear mention of instrument effects was detected in Adams (2000, p. 622); “Interviewing about the past is always problematic because of selective recall. However, there was enough similarity among different people’s responses to be able to make claims about what the workshops, family life, and the political and economic environment were like at the time”.

A total of 60 percent of the articles (39 out of 64) did not mention fallacies. In 15 percent of the articles (9 out of 64) a mention of fallacies was clearly supplied. For a quarter of the articles (16 out of 64) an unclear mention of fallacies was detected. Examples of unclear mention of fallacies were:

- Urkidi (2011, p. 568) who states “Slight differences may exist about the conceptualizations of community among actors from different scales and different strategic traditions, but here only the commonalities are analyzed”
- Davis (1994, p. 377) who states “if there is little consensus over what is meant by democratisation, claims about its origins can hardly be generalised, at least with the confidence implicit in much of the current literature on social movements and democracy”.

A clear mention of fallacies was detected in an article by de la Dehesa, (2007, p. 35) who states “[...] the analytic category of "social movement" can be deceptively unifying, obscuring organizational, strategic, and ideological differences as well as differential access to resources and relations of power (Melucci 1996; Rucht 1996; Chalmers, Martin, and Piester 1997) The concept of "field" offers one useful way to take this internal heterogeneity into account (Armstrong 2002)".

Almost two-thirds of the articles (41 out of 64) were not transparent regarding the conclusions reached. This includes six articles which either did not supply a CRQ or missed a link to the CRQ and 35 articles which supplied unclear answers to the CRQ. Furthermore, in 20 percent of the articles (14 out of 64) one or more conclusions emerged from an unspecified source. An example of a conclusion from an unspecified source was detected in Dosh (2006) who compared land invasion organizations in Lima and Quito. He asked what broader – non-neighborhood level factors – help explain citywide trends (p. 30) since neighborhood level factors according to him accounted for part of the variation. The use of an “original body of data collected in 2001–2002” (p. 50 note 2) partly consisting of “105 structured and open-ended interviews with past and present settlement leaders, as well as government officials, landowners, and other relevant parties” (ibid) lead to the conclusion that “Although neighborhood-level factors contribute to these differences, three national or citywide factors emerge as important explanations of metropolitan trends: public policy, local democratization, and geography and climate. However, while city-specific findings represent an important contribution, they are most notable for the questions they leave unanswered. [...] the conclusion that national and
metropolitan-level analysis can explain only a handful of outcomes in this context confirms the sustained importance of neighborhood-level factors." (p. 31).

Figure four shows a comparison of the numbers of aspects of transparency supplied of a reading that only takes into account the number of transparency aspects clearly supplied with a reading that counts the number of transparency aspects present also taking the number of unclearly supplied aspects into account. The majority of articles had between four out of seven and all seven aspects of transparency supplied. However, quite often articles had less than four out of seven aspects of transparency clearly supplied. The score on the transparency index ranged from 0.07 towards 0.93 (with mean 0.46 and standard deviation 0.23).

![Figure 4: "Loose reading" and "Strict reading" frequencies.](image)

Notes: * For a "strict reading" the clear aspects present were counted. ** For a "loose reading" both the unclear and clear aspects were counted.

### 5.3 Structure

Two-thirds of the articles (42 out of 64) were unstructured. That is, in 42 articles it was not possible to detect and link research questions with data collection methods and conclusions. Almost one-third of the articles (19 out of 64) were moderately structured. No more than 5 percent (3 out of 64 articles) was structured such that it would be possible to detect in the body text research questions, data collection methods and conclusions, to link them, and to distinguish between theory, data, analysis, results, and conclusions.

In almost half of the articles (30 out of 64) there was little influence of the data gathered on the conclusion reached. The same amount of articles had conclusions both influenced by theory and data collected. The remaining four articles had conclusions which were clearly based on the data collected.

The first quarter of the article in the majority of articles supplied both the first research question and the first conclusion (Figure five). However, one-third of the articles (21 out of 64) did not supply data collection methods in the body text. Quite often (13 articles out of 64) the
footnotes supplied data collection methods. In eight articles a data collection method was neither supplied in the body text nor in footnotes. In two articles only the abstract supplied a research question. Articles with data collection methods in the body text included them either in the first or second quarter of the article.

Figure 5: First shown location of the research question, data collection method, and conclusion.

A quarter of the articles (16 out of 64) followed the sequence of research question first, the data collection method second and conclusion third. A total of 40 percent of the articles (26 out of 64) followed six other sequences:

- Research question and conclusion → data collection method (11 articles).
- Research question → conclusion → data collection method (5 articles).
- Research question → data collection method and conclusion (4 articles).
- Research question and data collection method → conclusion (4 articles).
- Conclusion → research question → data collection method (2 articles).
- Conclusion → research question and data collection method (1 article).

In almost half of the articles (29 out of 64) the proportion of document relating to methods used (data collection plus sampling plus analysis) was lower than one percent. In half of the articles (32 out of 64) this percentage was between one and ten. In no more than five percent of the articles (3 out of 64) the proportion of document on methods was higher than 10 percent. The number of articles with conclusion proportion of document higher than 10 percent was almost four times higher (11 out of 64). Almost three quarters of the articles (47 out of 64) had less than one percent of the article on analysis method. A quarter of the articles (16 out of 64) had between one and ten percent of the article on analysis method. Almost nine out of ten articles (56 out of 64) had less than one percent of the article on sampling method and in one out of eighth articles, sampling method covered between one and 10 percent.

---

The included one article for which the research question, data collection method and conclusion were on the same page.
5.4 Coherence

For an assessment of coherence one article remained which supplied clear research question, clear data collection methods, analysis, sampling and answers to the central research questions. Therefore, for the assessment of coherence articles assessed were transparent to some extent, based on the related match rather than completely transparent. From the articles which both supplied clear central research questions and clear conclusions, two-thirds (11 out of 17) had an equal number of central research questions and conclusions which answered central research questions. The remaining one-third of the articles was divided between three articles which had more central research questions than there were conclusions and three articles which had more conclusions which answered a central research question than there were central research questions. Table three provides an example of an article (see Kaup, 2008) in which there was a mismatch between the number of central research questions and conclusions.

None of the 64 articles was completely coherent\(^{76}\) relating to the object of interest, object of research, and object of conclusion. This included five articles which supplied clear research questions, data collection methods, and conclusions. These five articles included three completely incoherent articles, one article which was for 25% coherent and one which was for 75% coherent. Overall, for all 64 articles, the minimum was 0% and the maximum was 75%, with a mean of 22% and a standard deviation equal to 19%.

From seven articles which clearly supplied all the data collection methods and conclusions, the article of Baletti et al. (2008) had a mismatch between the data collection methods and conclusions. While Baletti et al. (2008) include a reflexive account in their article, they did not reflect on the limits of their data collection method (unstructured interviews with an unclear number of leaders of the MST and two leaders of the LPM). The authors reflected (p.301) “While we recognize that the opinions of these figures are not universally shared by all actors with the movement, we believe these interviews not only highlight the mentalities of key figures within the LPM at the current conjuncture, but their words also reflect the trajectory the movement has taken in its short history”. Their article both used a convenience sample without a description of analysis for a real world descriptive conclusion and used unstructured interviews as a single data collection method to form a conclusion with a reach higher than experiences. The other six articles were coherent regarding the data collection methods and conclusions.

---

\(^{76}\) 1) AOI %community, AOR %community, AOC %community; 2) AOI %event, AOR %event, AOC %event; 3) AOI %household, AOR %household, AOC %household; 4) AOI %Latin America, AOR %Latin America, AOC %Latin America; 5) AOI %multiple events, AOR %multiple events, AOC %multiple events; 6) AOI %multiple Latin American countries, AOR %multiple Latin American countries, AOC %multiple Latin American countries; 7) AOI %multiple social movements, AOR %multiple social movements, AOC %multiple social movements; 8) AOI %municipality, AOR %municipality, AOC %municipality; 9) AOI %nation, AOR %nation, AOC %nation; 10) AOI %neighborhood, AOR %neighborhood, AOC %neighborhood; 11) AOI %other aggregate level, AOR %other aggregate level, AOC %other aggregate level; 12) AOI %social movement, AOR %social movement, AOC %social movement; 13) AOI %subgroup social movement, AOR %subgroup social movement, AOC %subgroup social movement; 14) IOI %individual, IOR %individual, IOC %individual. A percentage of coherence of objects of interest, objects of research, and objects of conclusion was computed in the following way using these 14 possible options used during data extraction: (total of 14 coherent / total of 14 present) * 100. Using this approach, data extracted regarding other aggregate level was not assessed. Also, it was not assessed what other codes occurred with individual objects of interest, research and conclusion. Last, multiple questions, methods, and conclusions were not taken into account.
Table 3: Incoherences in clear reporting of RQ, DCM and Conclusion

<table>
<thead>
<tr>
<th>DCM and Conclusion</th>
<th>RQ and Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Baletti et al., 2008)</td>
<td>(Kaup, 2008)</td>
</tr>
</tbody>
</table>

**Research question**

-In this paper, we analyze the political dynamics of the contemporary TPN through a comparison of two movements: the Brazilian Movimento Dos Trabalhadores Rurais Sem Terra (MST) and the South African Landless People’s Movement (LPM). *(p. 291)*

-"We do not attempt to provide a comprehensive explanation for the LPM's failure in this article. Rather, we explore the relationship between the LPM and the MST in an attempt to better understand the dynamics of the transnational peasant networks – and how these dynamics may have shaped the LPM's trajectory." *(p. 309)*

**Data collection**

-"[...] we describe the political and economic context of land distribution in Brazil and draw upon interviews conducted with MST activists in order to [...] we then describe the LPM from its [...] We analyze two interviews with critical movement leaders in order to develop and situate our analysis." *(p. 292-293)*

-"I use data from archival sources, interviews, and event observation. I obtained archival data [...] I conducted 45 open-ended interviews and observed several rallies, protests, and town hall meetings from [...]" *(p. 1735)*

**Conclusion**

-"In our comparison of the MST and the LPM, therefore, we argue that the latter is the 'late mobilizer', and as such it has experienced both advantages and disadvantages of engaging with the TPN. [...] Ultimately, the transfer of movement knowledge from the MST to the LPM may have worked against the long-term success of the latter." *(p. 292)*

-"We argue that the MST owed its success in Brazil to three main elements: [...] The MST then successfully translated its national success into a new peasant internationalism. The movement broadened its message and widened its support base throughout the 1990s." *(p. 292)*

-"As the MST became an important transnational actor, movement activists travelled to places like South Africa to support and advise rural activists there. Unfortunately, the MST's main strategy for mobilization – the land occupation – backfired in the South African context. The nearby example of racial tension in Zimbabwe made occupying land politically explosive in South Africa. Occupations were also difficult for the LPM to organize because the movement was unable to draw on the MST's other two elements of success: a leadership developed through grassroots experience and autonomy from civil society organizations and the state." *(p. 292)*

---

*This was an adequate rephrasing of the central research question previously identified.

**Based on sections in this article it was noticed that unstructured interviews and non-participant observation were conducted.

***Does not answer central research question. This conclusion is a partial answer (answers sub question).

****This conclusion was identified without certainty whether it was a conclusion or an implication for theory.

Two articles out of 23 articles which supplied clear conclusions had a mismatch between the study design and the highest reach of the conclusion. From these two articles, one used a case study design for a predictive conclusion and another used a cross-sectional design for a predictive conclusion. None of the articles had a mismatch on the ground that the article used a case study design for a prescriptive conclusion.

A quarter of the articles (10 out of 41) which supplied clear research questions had a mismatch between the type of research question and the study design. These articles were
divided between five articles which used a case study for longitudinal questions and five articles which used a case study for causal questions (see also Kaup, 2008).

One out of 17 articles which both supplied clear research questions and conclusion had a mismatch\(^\text{77}\) between the reach of research questions and reach of conclusions. In the article with a mismatch, the highest reach of the research question (experiences) was lower than the highest reach of the conclusion (real world descriptive). The 16 coherent articles were divided between 15 articles in which the highest reach of the research question was equal to the highest reach of the conclusion and one article in which the highest reach of the research question was higher than the highest reach of the conclusion.

All seven articles assessed for coherence between types of research questions and types of conclusions\(^\text{78}\) were coherent. Because the number of research questions and conclusions were not similar, it was not possible to assess whether there was coherence between the types of the research questions and types of conclusions. Out of 17 articles which supplied both clear research questions and conclusions ten articles had an unequal amount of research questions and conclusions and the remaining seven articles were coherent regarding the types of questions and conclusions.

---

\(^{77}\) Reaches of research question and conclusion were ordered from low towards high; research context, experiences, real world descriptive, real world in the past, real world prescriptive and real world predictive.

\(^{78}\) Coherence between types were cases where the number of research questions equalled the number of conclusions and the sum of all research question types minus the conclusion types was zero, i.e. \((\text{total } RQ \text{ descriptive} - \text{total } C \text{ descriptive}) + (\text{total } RQ \text{ correlational} - \text{total } C \text{ correlational}) + (\text{total } RQ \text{ longitudinal} - \text{total } C \text{ longitudinal}) + (\text{total } RQ \text{ causal} - \text{total } C \text{ causal}) = 0\).
6. Discussion

The attempt to apply the “simple” BS checklist by Tamás and Kampen (2013) to research articles on social movements in Latin America turned out not to be a simple task. Articles severely lacked transparency and structure, which made the assessment of internal coherence inconsistent. To operate on the level of internal coherence of research reports assumes that aspects necessary for the assessment of coherence are clearly reported. The question “do the conclusions answer the central research questions?” implicitly assumes that (1) central research questions and conclusions are explicitly reported (transparency) and (2) that a conclusion is reported after a central research question (structure). Interpretation of aspects in the reported research has to be made in case of non-explicit reporting. Systematic quality assessments of a number of articles based on interpretations of reports are not systematic.

Therefore, the results of this systematic review of the coherence of research articles should not be considered as rigorous results. The low number of articles which remained for the assessment of coherence does however reflect the lack of transparency in the research articles included in this review. The results of the assessment of coherence first suggest that next to a lack of transparency there is also some reason for concern regarding coherence between the number of research questions and the number of conclusions (although this overlaps with structuredness). Second, the results of the assessment of coherence show incoherence between the study design and the type of research questions.

The result that none of the articles was completely coherent regarding the object of interest, object of research and object of conclusion is an indication rather than a rigorous result. The assessment was based on interpretation of unclear reporting and assessment of individual coherence aspects overlapped with other individual assessments such as the number of questions and number of conclusions. Also, the criteria that all objects of interest, objects of research and objects of conclusions would have to match was a rather strict one for research on social movements since it is impossible to study a complete social movement. Similarly the result of the influence of data on the conclusions should not be considered as rigorous results, since articles lacked a separation between theory, data collection methods, data analysis, results, conclusions, and implications.

A strict application of the “BS checklist” by Tamás and Kampen (2013) to individual research articles would have led to a high number of articles being excluded from this review on the ground that it was not possible to start or continue a "BS assessment". Critics would probably first bring up the point that inadequate reporting does not make the actual research conducted bullshit. While this point is certainly true, research articles are often the only site to evaluate research and consumers of research who are highly dependent of the results of the research cannot accept research articles at face value. If the reporting is inadequate a decision on the usefulness of the knowledge generated cannot be made by consumers of research articles.

Another point critics may raise to checklist style BS detection is that each individual article has to be evaluated on its own merits and that it is up to consumers of research to assign relevance. Using predefined criteria may not take into account that "progress has been made in developing new tools, methods, and approaches to social movements" (van Stekelenburg & Roggeband, 2013) and that articles using such new tools are relevant for consumers of research. Articles from the journal Latin American Perspectives were included in the current review, whereas the journal indicates that articles are presenting a story, a point of view, rather than empirical research. However, even for consumers of a perspective it is necessary to observe
where the perspective comes from. The position that each individual study should be evaluated on its own merits disallows any method of establishing legitimacy, validity, credibility etc. There is no reason why qualitative (social movement) research cannot ensure rigor (Krefting, 1991; Popay & Williams, 1998; Seale & Silverman, 1997) and consumers of research articles need to be able to see how the rigor has been achieved before they can accept the findings of a study. Information about how (new) methods to ensure rigor are used is crucial for progress in qualitative (social movement) research.

A last point critics may probably raise is that any checklist to assess research implicitly includes an idea of what research is and what it should be. In the transparency index developed for example there are reporting criteria for analysis, sampling, and sample sizes, whereas in anthropological research it may be the case that these aspects are not necessarily given a high priority. Even if sampling and sample sizes would not be of a high priority in anthropological research, there is not a reason as far as the author knows not to report how access was gained towards the population of study. While it may be the case that sampling and sample sizes are not a high priority for some types of research, it needs to be mentioned that other checklists are considerably stricter. In this systematic review questions stated without a question mark were also clear central research question, although such questions still implied some interpretation. Also, for data collection methods, the transparency index included unclear mention, and for sampling and analysis it was simply assessed whether they were reported or missing without going into the details of appropriateness. Last, in the assessment of coherence a conclusion with the reach experiences based on unstructured interviews were also part of a coherent article. A systematic review has been chosen for transparency reasons and it is expected by the author that attempts at replications find lower ratings on the transparency index.

Where some scholars may see prescriptive checklists as a threat to qualitative (social movement) research, the argument made here is that the checklist as used in this systematic review, rather than being prescriptive is intended to assess articles as reported and research based on the questions asked rather than epistemological preferences. Use of the BS checklist developed by Tamás and Kampen (2013), and the transparency index from the current thesis, is intended to be a caution against other threats, such as the development of real world descriptive theories based on an over-reliance of (cherry-picked) verbal representations.

This systematic review provides a comment for a change of use of the BS checklist. It proposes a quick scan through articles for transparency and structure before application of the “bullshit detector”. It is necessary, because non-presence, or unclear presence, of either (a) central research questions (b) data collection- (c) sampling- and (d) analysis- methods; (e) study design; (f) results; (g) involved theory or (h) conclusions, interrupts the conduct of systematic quality or coherence assessments. This is a simple scan and involves a count of the mention of a number of determinants of quality and provides a first indication whether important aspects of the research process can be separated.

A count of clear mention of determinants of quality could also make going through a number of articles for quality assessment more systematic. At the moment different checklists mix questions of transparency and adequacy and include a possibility for evaluators to add written comments for questions in the checklist. In this form, a final decision on quality remains non-systematic in the sense that step-by-step guidelines for evaluators and standard measurement for specific items in checklists are missing.

Rating articles for transparency could make assessments systematic because:
• By using a rating system different studies are handled the same way based on interpretability
• Differences between evaluators of research can easily be identified.
• A minimum total score for transparency can be set as standard for inclusion in policy supporting systematic reviews.

This systematic review has noticed that a call for “more standardization of data collection and reporting” (Bert Klandermans, Staggenborg, & Tarrow, 2002, p. 333) has not yet been taken up for research articles on social movements in Latin America. Standardization of reporting is necessary for social movement research not to be self-excluding from systematic reviews. If research articles are not transparent and reflexive, there remains reason for consumers of research to see the work as “bullshit” or “cargo cult science” (Feynman, 1974). Smelling BS is one thing, systematically detecting it is another. Identifying whether an article has a poker face helps.
7. Limitations

A first limit relates to incoherence between the literature review on which the data extraction form was based and literature on the topic of this thesis. The literature used in order to develop the data extraction form and to operationalise the concepts of transparency, structure and coherence relates to quality in qualitative research in the field of health care. The field of health care differs to a large extent from the area of social movement research, for example with respect to the aims, design, context, and audience of research. While health care research is in the majority of cases intended to support policy, in social movement research articles also have other goals such as suggesting possible hypotheses for colleague researchers. While it could be argued that the incoherence between literature used and topic has influenced the outcome – for example by including “all sample sizes supplied” as a criteria in the transparency index – there are no quality criteria designed specifically for the area of social movement research and in the execution of this systematic review a lenient approach has been taken to the included articles.

Second, there is a concern relating to reliability of the data extraction. It may be the case that at certain point frustration started being measured rather than specific items in the data extraction form since often from the start when extracting data regarding the research questions it proved difficult to interpret what questions were asked by authors when they entered the field. Internal validity was dealt with by extracting data in several rounds within a cohort, rather sample a number of articles and directly apply the full data extraction to articles. While it was an option to measure the time spent coding each article as a measure of reliable data extraction, the choice to sample in cohorts and extract data in steps was seen as more practical and efficient. The chosen option was not compatible with measuring time per article needed.

Third, there are uncertainties regarding the search and sample strategy. The first uncertainty regards the choice to search in Web of Science rather than Scopus, which may have influenced the total population articles identified. A reference categorized as an article in Web of science could be categorized as a review article in Scopus and abstracts were often not clear whether the article was primary research or not. Thus, classification also depended on categorization in Web of Science. In order to get a measurement of validity for classification, from 93 articles classified as secondary research, a random sample of 18 articles was taken. Out of these 18 articles, based on a scan of the full article, two were primary research and for two articles it was unclear whether the article was primary or secondary research. An indication would thus be that between 10 and 20 percent of articles were misclassified as secondary research. However, this misclassification would not have influenced the extraction of data, but the total number of 219 articles identified as population. Articles which were clearly secondary research were excluded during the data extraction. A second uncertainty regarding the search strategy was the possibility that Spanish articles were being artificially excluded when developing the syntax. During the development of the search syntax, lists of journals in which the articles found through Web of Science and Scopus were compared with electronically accessible social science journals of the Wageningen University Library. Although in the majority of cases there was no access to Spanish journals, there were no serious access problems. There was access to the journals including those which published most articles out of the search results. The journals to which there was no access mainly published a single article. A last problem relates to counting errors during sampling which makes a proof of the randomness of the sample difficult. The sample was random as articles were not read before being sampled and they were ordered alphabetically rather than a certain topic or journal. It is important
nevertheless, that care is exercised when making statements about the larger population of research articles on social movements in Latin America.

Fourth, the proportion of documents relating to the methods used was limited to data collection methods, sampling methods, and analysis methods. Operationalisation of concepts and justifications for operationalisation were not included and neither were contexts which may have influenced the decision to select specific cases in case studies. This decision to focus on technical aspects of methods used for the methods proportion of documents needs to be taken into account when interpreting the outcome of the percentage of the article on methods used.

Last, there were a number of difficulties encountered in efficient use of the Atlas.ti data analysis software since the program is not specifically developed for systematic reviews. For example, Atlas.ti does not support dealing with bibliographic data. Therefore, data on citations per year, journal title and year of publication were not exported from Atlas.ti towards SPSS but were manually dealt with in Excel. Another difficulty was the export from Atlas.ti towards SPSS before the analysis phase. Also, articles were to a large extent coded on individual research question, data collection method and conclusion level rather than on an article level. In SPSS the output was intended to be on an article level. Dealing with these procedural difficulties was time consuming. Manual input in SPSS next to the use of Atlas.ti could have been more efficient. However, what was encountered in this research was not expected beforehand, and the chosen strategy provided the flexibility needed to fix for unreliable data extraction which would not be possible when data would be inserted manually in SPSS next to Atlas.ti.
8. Conclusion and Outlook

The first question this systematic review asked was: *To what extent are empirical social science articles on Latin American social movements transparent and structured?* The articles on social movements in Latin America included in this systematic review were neither transparent nor structured sufficiently. The articles were transparent to the extent that it was possible to identify research questions, data collection methods and conclusions. However, these three aspects were often unclear. Analysis methods, sampling methods and a mention of fallacies or instrument effects were often not supplied at all. Articles did not go further regarding structure than the possibility to identify research questions, data collection methods, and conclusions, i.e. the overlap with transparency. Making a link between these aspects, together with a separation between the research process and the research results was in the majority of cases not possible.

The second question asked was: *to what extent are transparent articles also coherent?* A reliable conclusion on the coherence of research articles cannot be made since a small number of articles remained after excluding non-transparent articles for each assessment of coherence. However, there may be reason for concern regarding incoherence between the study designs and the type of research questions asked, and incoherence between the number of conclusions and number of central research questions.

Together with the editors of one of the scarce books on research methods for social movement research (Bert Klandermans, et al., 2002, p. 333) this review calls for “greater clarity in reporting the procedures used, such as questions asked in interviews and methods of coding and analysis”. A radical starting point in the effort to standardize reporting would be to join other scholars and editors in developing “new guidelines capable of reflecting a distinct understanding of what constitutes legitimate knowledge and of how such knowledge is to be reported” (Zeller & Farmer, 1999, p. 17). However, for the moment, it is proposed that (Latin American) social movement researchers use available guidelines for reporting their research, such as COREQ (Tong, et al., 2007) or an older guide developed by Knafl and Howard (1984). For editors and reviewers of research it is proposed that they use indicators such as the developed transparency index in order to publish transparent articles or include adequately reported transparent articles in reviews.
References


Carroll, C., Booth, A., & Lloyd-Jones, M. (2012). Should we exclude inadequately reported studies from qualitative systematic reviews? an evaluation of sensitivity analyses in two case study reviews. *Qualitative Health Research, 22*(10), 1425-1434.


Denzin, N. K. (2009). The elephant in the living room: or extending the conversation about the politics of evidence. *Qualitative Research, 9*(2), 139-160.


Appendix 1: Data extraction form

This data extraction form was used when coding each individual article in Atlas.ti. Between brackets the codes used are written in full length. In practice abbreviated codes were used.

* Answered for each individual research question; **Answered for each individual data collection method; ***Answered for each individual conclusion

**ID:** (Author; Year; Title; Journal; Citations per year)

**Structuredness:** ((Low; Weak; Moderate; High)

**Body text:** (All text including footnotes – excluding title, abstract, acknowledgement and references)

**Number of pages with body text:** (Open)

**Study design:** (Experimental, Cross-sectional, Longitudinal, Case study; Case study-longitudinal, Case study-multiple, Case study-comparative)

**Proportion of document on research questions:** (Word count using three codes, one for research gaps, one for places with article structure and one for stated questions)

**First shown location of the research question:** (abstract, open)

* **Number of research question:** (Open)

* **Research question origin:** (Central research question; Sub question; Rephrasing central research question-adequate; Rephrasing central research question-inadequate; New)

* **Research question type:** (Causal; Correlational-directional; Correlational-non-directional; Longitudinal; Descriptive)

* **Research question clarity:** (Stated with?; Stated without; Implicit; Unclear)

* **Research question reach:** (Experiences; Research context; Real world-description; Real world- predictive; Real world-prescriptive; Real world-past; Real world-interpretation; Real world-imagination)

* **Individual object(s) of interest:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

* **Individual other aggregate level:** (Open)

* **Aggregate object(s) of interest:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

* **Aggregate object of interest-other aggregate level:** (Open)

**Proportion of document on methods used:** (word count using three codes ‘data collection method’ ‘how sampling’ and ‘how analysis’)

**First shown location of the data collection method(s):** (Open; Footnotes; None)

**Reporting sample size(s):** (All reported; Incompletely reported; Unspecified; No mention sampling)

**Mention of ecological fallacy or the exception fallacy:**(Yes; Overstatement; No)

**Mention effect of instrument used:** (Yes; Overstatement; No)

**Number of data collection method:** (Open)
**Data collection method(s):** (Biological; Interview (includes structured/unstructured/unknown); observation (includes participant/non-participant/unknown); Document review; Understatements and overstatements (possible for all of the above))

**Sampling type:** (Random; Convenience sample; Purposive sample; Convenience and purposive; Unspecified sample; No mention sampling)

**Individual object(s) of research:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

**Individual object of research other aggregate level:** (Open)

**Aggregate object(s) of research:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

**Aggregate object of research-other aggregate level:** (Open)

**Proportion of document on conclusions:** (Word count using one code 'conclusion proportion of document')

**Influence of theory on all the conclusions:** (Little; Empirical and Theory; High)

**First shown location of a conclusion in the body text:** (Open)

**Conclusion number:** (Open)

**Conclusion link with research question:** (Answers central research question; Answers sub question; No link with a research question; Answers central research question-overstatement)

**Conclusion source:** (Theory, Empirical, Empirical and theory, Unspecified source)

**Conclusion type:** (Causal; Correlational-directional; Correlational-non-directional; Longitudinal; Descriptive)

**Conclusion reach:** (Experiences; Research context; Real world-description; Real world-predictive; Real world-prescriptive; Real world-past; Real world-interpretation; Real world-imagination)

**Individual object(s) of conclusion:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

**Individual object of conclusion-other aggregate level:** (Open)

**Aggregate object(s) of conclusion:** (Individual; Household; Neighborhood; Municipality; Community; Society; Nation; Multiple (Latin American) countries; Latin America; Subgroups of social movement; Social movement; Multiple social movements; Event; Multiple events; Other aggregate level)

**Aggregate object of conclusion other aggregate level:** (Open)

**Proportion of the document on implications:** (Policy recommendation; Social movement recommendation; Concept use; Future research; Limitations; Theory)
Appendix 2: Articles included in the review (N=64)


Appendix 3: Articles classified as secondary research (N=93)


76 Sanchez, R. M. M. (2012). Building knowledge from the margins: information, knowledge and social movements. [Article]. Transinformacao, 24(1), 61-64.


Appendix 4: Web of Science result (N=549)


Markowitz, L., & Tice, K. W. (2002). Paradoxes of professionalization - Parallel dilemmas in women’s organizations in the Americas. [Article]. Gender & Society, 16(6), 941-958.


