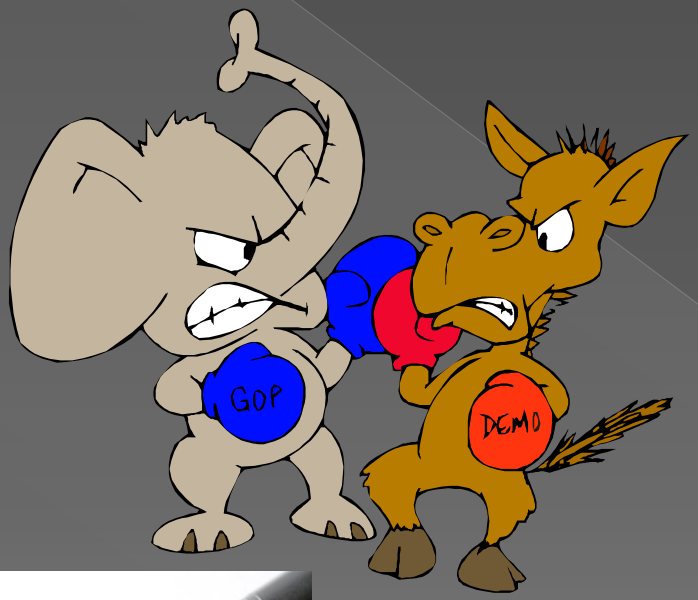
Critical Appraisal of Qualitative Research

Choosing an instrument (part 1) and Choosing wether to appraise studies or findings of studies (part 2)

Karin Hannes
Methodology of Educational Sciences Research Centre
KU Leuven

The Aspirine Case

To appraise or not to appraise



- The more you appraise, the lesser the chance to end up with flawed results. The main criterion is **quality**!
- The more you appraise, the more it stifles creativity. The main criterion is **relevance**!

PART 1:
How to appraise?
Which instrument to choose?

Your options

Expert judgement
Critical appraisal instrument or framework

MAKING SENSE OF THE MYRIAD OF CRITICAL APPRAISAL INSTRUMENTS

2 points of view

- **A pragmatic point of view**
- **An epistemological point of view**

Which instrument?

- **A pragmatist point of view:**
  - The choice for using certain critical appraisal instruments (CAI) should be based on the 'utility' and 'fit for purpose' of the CAFs for the studies to be included in the reviews.
  - Reviewers should select CAIs that are suitable for the retrieved original studies.

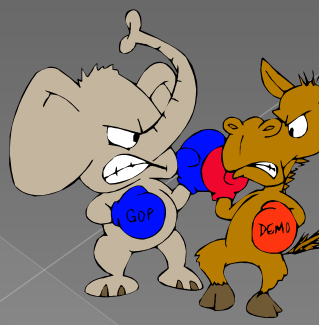
Which instrument?

- **A pragmatist point of view:**
  - If you are appraising a transformative-emancipatory article you would use a CAI including transformative-emancipatory criteria:
    - Do the authors openly reference a problem in a community of concern?
    - Do the authors openly declare a theoretical lens?
    - Were the research questions or purposes written with an advocacy stance?
    - Did the literature review include discussions of diversity and oppression?
    - Did the authors discuss appropriate labeling of the participants?
    - Did data collection and outcomes benefit the community?
    - Were the stakeholders involved in the research project?
    - Did the results elucidate power relationships?
    - Did the results facilitate social change, and were the stakeholders empowered as a result of the research process?
    - Did the authors explicitly state their use of a transformative framework?

(Saini & Shlonsky, 2012, pp. 133-135; Sweetman, Badiee, & Creswell, 2010, pp. 442-443)

Which instrument?

- **An epistemological point of view:**
  - The choice for the CAIs can, implicitly or explicitly, be influenced by the philosophical stance of the researchers doing the reviews.
  - Based on their philosophical stance:
    - They can 'value' some CAIs, and some quality criteria included in CAIs, more than others.
    - They could value generic appraisal tools above self-composed or design-specific appraisal tools.

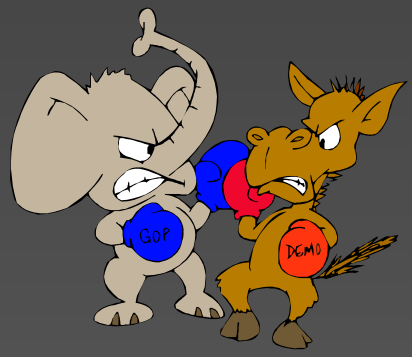


Which instrument?

- **The easy way out:**
  - a universally accepted CAI → Not realistic

  What else?
  - a 'pick-and-choose' CAI: the researcher composes a set of criteria (from a larger pool of criteria) and account for this set based on his philosophical stance.
  - provide a set of criteria that should minimally be addressed in each primary study and add criteria to it, either for design or for epistemological purposes.

Which instrument?



An epistemological point of view:

- **a *realist* researcher** who highly values the validity of primary studies would prefer a CAI that is sensitive to aspects of validity, including criteria such as '***all statements should be well-grounded in the data***' and '**the impact of the investigator on the study results should be reported**'.
- Validity and researcher bias should be evaluated → Some qualitative studies are more rigorous than others.
- **an *interpretivist* researcher** would subscribe to the argument that the impact of the researcher on the research is inherent to the way qualitative research is conducted. They may prefer a CAI evaluating issues such as '**thick description**' and '**the innovative nature' or 'value for practice'** of the findings.
- Used to state: Epistemological and ontological assumptions of quantitative and qualitative research are incompatible → It is inappropriate to use such measures -> Now: Translation of terms.

Which instrument?

- Hannes, Lockwood & Pearson ('10)
  Appraisal Instruments' Ability to Assess Validity in Qualitative Research
- Selection of appraisal instruments:
  - Used in recently published QES (2005-2008)
  - Online available and ready to use
  - Broadly applicable to different qualitative research designs
  - Developed and supported by an organisation/institute/consortium or a context other than individual, academic interest.

Which criteria are used to evaluate the quality of a study?

- Three instruments fit the inclusion criteria:
  - Joanna Briggs Institute-Tool
    http://www.joannabriggs.edu.au/cqrmg/tools_3.html
  - Critical Appraisal Skills Programme-Tool
    http://www.phru.nhs.uk/Doc_Links/Qualitative%20Appraisal%20Tool.pdf
  - Evaluation Tool for Qualitative Studies
    http://www.fhsc.salford.ac.uk/hcprdu/tools/qualitative.htm
- To facilitate comparison:
  - Criteria grouped under 11 headings
  - Cross-comparison of the criteria (lead criteria from the instruments)

Criterion	JBI Tool	CASP Tool
	There is congruity between:	Screening questions: Was there a clear statement of the aims? Is a qualitative methodology appropriate?
Theoretical framework	The stated philosophical perspective and the research methodology	
Appropriateness of research design	The research methodology and the research question or objectives	Was the research design appropriate to address the aims of the research?
Data collection	The research methodology and the methods used to collect data	Was the recruitment strategy appropriate to the aims of the research? Were the data collected in a way that addressed the research issue?
Data analysis	The research methodology and the representation and analysis of data	Was the data analysis sufficiently rigorous?
Findings	The research methodology and the interpretation of results	Is there a clear statement of findings?
Context	There is a statement locating the researcher culturally	
Impact of investigator	The influence of the researcher on the research, and vice versa, is clear	Has the relationship between researchers and participants been adequately considered?
Believability	Participants, and their voices, are heard	
Ethics	The research is ethical according to current criteria, or there is evidence of ethical approval by an appropriate body	Have ethical issues been taken into consideration?
Evaluation/ outcome	Conclusions drawn in the research report do appear to flow from the analysis, or interpretation, of the data	
Value and implications of research		How valuable is the research?

Which criteria are used to evaluate the quality of a study?

Criterion	JBI	CASP
		Screening Q
Theoretical framework		NO
Appropriateness design		
Data collection procedure	Both instruments have focussed on the accuracy of the audit trail = Quality of reporting.	
Data-analysis procedure		
Findings		
Context		NO
Impact of investigator		
Believability		NO
Ethics		
Evaluation/Outcome		NO
Value/Implication Research	NO	

Which instrument?

- Realist researcher: We need to know
  - whether the set of arguments or the conclusion derived from a study necessarily follows from the premises.
  - whether it is well grounded in logic or truth.
  - whether it accurately reflects the concepts, the ideas that it is intended to measure.

(What is) validity (?)
→ operationalisation using Maxwell's framework

Validity as the main criterion

- Maxwell's deconstruction of the concept validity (1992)
  - Descriptive validity
  - Interpretative validity
  - Theoretical validity
  - Generalisibility (external validity)
  - Evaluative validity

Maxwell	Definition	Techniques
Descriptive validity	The degree to which descriptive information such as events, subjects, setting, time, place are accurately reported (facts).	Methods- & Investigator triangulation → allows for cross-checking of observations
Interpretative validity	The degree to which participants' viewpoints, thoughts, intentions, and experiences are accurately understood and reported by the researcher.	Display of citations, excerpts, use of multiple analysts (inter-rater agreements), self-reflection of the researcher, (member checking)
Theoretical validity	The degree to which a theory or theoretical explanation informing or developed from a research study fits the data and is therefore credible/defensible.	Persistent observation → stable patterns, deviant or disconfirming cases, multiple working hypotheses, theory triangulation, pattern matching
Generalisa-bility (external validity)	The degree to which findings can be extended to other persons, times or settings than those directly studied.	Demographics, contextual background information, thick description, replication logic
Evaluative validity	The degree to which an evaluative critic is applied to the object of study (as part of the researcher's reflexivity)	Ethics?, Clarifying the links between conclusions and other parts of the research process? Application of an evaluative framework

What is the extent to which the different instruments assess validity in their lead criteria?

Table 2. Types of Validity Addressed in the Critical Appraisal Instruments

Types of Validity	Description	Criteria	Appraisal Instruments
Descriptive validity	The degree to which descriptive information such as events, subjects, setting, time, and places are accurately reported	Impact of investigator	Evaluated in JBI, CASP, & ETQS
		Context	JBI & ETQS
Interpretive validity	The degree to which participants' viewpoints, thoughts, intentions, and experiences are accurately understood and reported by the qualitative researcher	Believability	Evaluated in JBI & ETQS
Theoretical validity	The degree to which a theory or theoretical explanation informing or developed from a research study fits the data and is, therefore, credible and defensible	Theoretical framework	Evaluated in JBI & ETQS
Generalizability	The degree to which findings can be extended to other persons, times, or settings than those directly studied	Value and implications of research	Evaluated in CASP & ETQS
Evaluative validity	The degree to which an evaluative framework or critique is applied to the object of study	Evaluation/outcome	Evaluated in JBI & ETQS

Note. JBI = Joanna Briggs Institute; CASP = Critical Appraisal Skills Program; ETQS = Evaluation Tool for Qualitative Studies

What is the extent to which the different instruments assess validity?

- The most commonly used instrument 'CASP', is the least sensitive to aspects of validity (findings based on screening the main headings). It does not address interpretive nor theoretical validity or context as a lead criterion.
  - The theoretical position and the background of a researcher has a direct impact on the interpretation of the findings.
  - Statements that have no clear link to excerpts are at risk of not being grounded in the data.
  - Therefore, they should be **LEAD CRITERIA** in a critical appraisal instrument!

What is the extent to which the different instruments assess validity?

- The study is limited by
  - Its focus on the **lead criteria** of the instrument. Some of the subheadings of CASP do address issues of e.g. interpretive validity and some issues are not addressed in the JBI-tool, e.g. sampling procedures.
  - Its focus on **validity** as an evaluation criterion.

Which instrument?

- Would it help to position yourself on the epistemological spectrum to facilitate the choice of a CAI?
- Are there fundamental differences between CAIs?
- Do we need to give our choice of CAI considerable thought?
- Could CAIs assist us in evaluating the (methodological) quality of a study?
- Do CAIs help us to establish rigor in qualitative research?

To conclude this first part

- In evaluating validity at the end of a study (post hoc), rather than focusing on processes of verification during the study we run the risk of missing serious threats to validity until it is too late to correct them.
- I argue for the use of verification techniques for validity as a means for obtaining rigor.
- Basic qualitative researchers
  - should be motivated to adopt techniques that improve validity
  - Should be guided in how to report qualitative research in order to facilitate critical appraisal
- Most importantly, the CAI should be applied consistently across studies.

PART 2:
Weighing evidence
Appraising Studies versus appraising findings

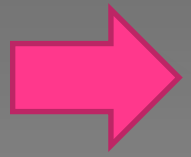
1. Sensitivity analyses
2. Levels of evidence assigned to findings
3. Frequency of themes combined with quality appraisal.

Weighing the impact of studies: sensitivity analysis

- Sensitivity analysis involves testing how sensitive the review findings are to the inclusion and exclusion of studies of different quality.
- Question: What would happen to the results if all studies below a certain quality threshold would systematically be excluded?

Weighing the impact of studies: sensitivity analysis

- Carroll, Booth & Lloyd-Jones ('12)
- Examining the effect of exclusion of studies on the basis of the adequacy of **reporting of methods** (adequacy with which procedural elements are described in a study) for 2 reviews.
  - The question and study design
  - The selection of participants
  - Methods of data collection
  - Methods of analysis



Dichotomization of studies:
Adequately (>2)
Inadequately (<2) reported

Weighing the impact of studies: sensitivity analysis

- 9/19 studies were judged to be inadequately reported for both reviews.
- **REVIEW 1: sexual health studies**
  - No single principal theme was completely dependent on data from inadequately reported studies!
  - No data emerged as exclusive findings from the inadequately reported studies.
  - With the exception of two viewpoints from lower quality studies, all instances of dissonance, richness or complexity for each theme emerged from on or more of the adequately reported studies.
- Consistent with studies from other authors having conducted sensititivity analyses (Noyes and Popay 2007, Thomas and Harden 2008)

Weighing the impact of studies: sensitivity analysis

- **REVIEW 2: Online learning**
  - Overall, data derived from inadequately reported studies did little to supplement data from adequately reported studies.
  - Some of the richness was generated from inadequately reported studies.
  - Excluding them would have resulted in the loss of valuable data on one particular subgroup (nurses).
  - As a consequence differences between doctors and nurses might have been conceiled.

Weighing the evidence: levels of evidence for findings

- JBI levels of evidence

Line of argument:

Author statements can be considered unequivocal, credible or unsupported, based on how well they are supported with excerpts from the data collection phase.

- **Unequivocal:** Where the theme or metaphor is unequivocally supported by direct quotes from the research participants. There is a clear relationship between the author's theme or metaphor and the participants' expressed experiences.
- **Credible:** where the participants' expressed experiences are less clearly related to the author's theme or metaphor and the author has extended beyond the expressed experiences of the participants based on the participant quotes that have been used.
- **Unsupported:** where there is no relationship between the expressed experiences of the participants and the author's themes or metaphors, then it is clear that the author is generating findings that are unsubstantiated by the participants.

The author of a review chooses whether or not to include unsupported and credible findings.

Weighing the evidence: Frequencies

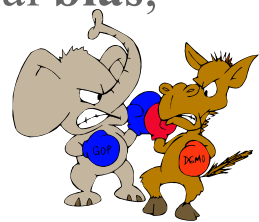
- Boeije en Van Wesel ('11)
  - Counting the **frequency** of a theme in an included article
  - Combine the frequency with the weight of quality appraisal done by **expert judgement (EJ)**
  - Combine the frequency with the weight of quality appraisal done by **checklists (CA)**

The evidence for high frequency themes increases, while for low frequency themes it declines!
The direction of the change is the same for EJ and CA, CA has a better differentiating ability.

Average score of 6.8 and Inter rater reliability=0.88

Average score of 6.9 and Inter rater reliability of 0.94

•Lower scores for criteria related to **validity** of the study, reporting of potential **bias**, **contextual** info in order to evaluate transferability of the findings.
•Some articles scored low on the checklist but passed the expert judgement based on **significance of the findings**!



Weighing the evidence: Frequencies

- Boeije & Van Wesel ('11)

Line of argument:

- When a topic is frequently studied in a methodologically sound way, there is strong evidence for the value of the findings.
- Not all studies are of equal methodological quality and this should be accounted for when integrating findings.
- Limitation: Working with a summary score may conceal errors that can be considered fatal.

To conclude…

- Checklists may only capture what has been reported (but reviewers can dig deeper ☺).
- Excluding findings might be a valuable alternative for excluding studies.
- To validate is to investigate, to check, to question, and to theorize. All of these activities are integral components of qualitative inquiry that insure rigor (Morse, 2002).
- The process of inquiry is where the real verification happens.

Conflicts of interest:
I am a Cochrane knight and a Campbell light

I might substantially have been brainwashed in the 'risk of bias' discourse, beyond my personal control.

Karin.hannes@ppw.kuleuven.be