The design and evaluation of a critical appraisal tool for qualitative and quantitative health research

Submitted by

Michael Crowe

MIT – National University of Ireland, Galway
BSc (Mgmt) – Trinity College Dublin, Ireland
ADMT – Dublin Institute of Technology, Ireland

in

May 2011

for the degree of

Doctor of Philosophy

in the school of

Public Health, Tropical Medicine and Rehabilitation Sciences, James Cook University

Supervisor  Prof Lorraine Sheppard
Associate Supervisor  Dr Alistair Campbell
Declarations

RELEASE

Except where otherwise acknowledge herein, this thesis is the author’s own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education.

In accordance with James Cook University’s Intellectual Property Policy (2010) all intellectual property created by the author in the course of undertaking this PhD project belongs to the author and the author owns copyright to this thesis.

COPYRIGHT

Every reasonable effort has been made to gain permission from and acknowledge the owners of copyright material. Any copyright owner who has been omitted or incorrectly acknowledged can contact the author to remedy this situation. Extracts of copyright permissions are included in Appendix A.

ETHICS

The research presented and reported in this thesis was conducted within the guidelines for research ethics outlined in the National Statement on Ethical Conduct in Human Research (2007), the Australian Code for the Responsible Conduct of Research (2007), and James Cook University’s Code for the Responsible Conduct of
Declarations

*Research* (2009). The proposed research received authorisation from James Cook University’s Human Research Ethics Committee (approval number H3415). A copy of the approval is reproduced in Appendix B.

**CONTRIBUTION OF OTHERS**

The author was awarded a JCU scholarship for the duration of their study. No other financial assistance was received. The author received no substantial research design, statistical, data analysis, or technical assistance apart from that duly provided by a PhD supervisor or associate supervisor. Ms Margaret Bowden provided professional editorial assistance which was limited to Standards D (language and illustrations) and E (completeness and consistency) of the *Australian Standards for Editing Practice* (Council of Australian Societies of Editors, 2001) in line with James Cook University’s *Proof-Reading and Editing of Theses and Dissertations* (2010) policy.

**INCORPORATED PUBLISHED WORKS**

The following articles by the author have been published or are under review for publication, and form an integral part of this thesis. Chapters where these articles appear are indicated. Where articles were co-authored, this was in co-operation with the PhD candidate’s supervisor or associate supervisor. Co-authorship contributions were towards the concept, drafting and final approval of the published work, as duly provided by a PhD supervisor or associate supervisor. Articles that were published are reproduced in Appendix C.

*Chapter 2 – Crowe, M., & Sheppard, L. (2010). Qualitative and quantitative research designs are more similar than different [Invited editorial]. Internet Journal of Allied Health Sciences and Practice, 8(4). Retrieved from http://ijahsp.nova.edu/*


By signing these declarations, the author confirms to the best of their knowledge that the statements are accurate and true.

6 May 2011

Michael Crowe Dated
Acknowledgements

Charlie and Dad, who were here for the beginning but not at the end.

Thanks to friends and colleagues at JCU for their encouragement throughout. Also, thanks to the participants in the research who volunteered their spare time to read and appraise papers.

My family and Mum, who think I’m a kind of ‘eejit’ for doing this in the first place. Even so, they have supported me all the way.

Lorraine Sheppard, generous, encouraging, invaluable, without your help I would still be thinking about doing a PhD rather than finishing. And Alistair Campbell, for helping me through decisions on design and analysis.

Most of all, my gratitude goes to Anne. Why we were doing PhDs at the same time is still a mystery. Your love, support, guidance, tenacity, and ability to cut through the bulldust are inspirational.
Abstract

OBJECTIVE
To design and evaluate a critical appraisal tool (CAT) that can assess the research methods used in a broad range of qualitative and quantitative health research papers; has the depth to fully assess these research papers; has an appropriate scoring system; and has validity and reliability data available to evaluate the scores obtained by the tool.

Critical appraisal is defined here as the impartial assessment of one or more research papers to determine their strengths, weaknesses and benefits.

STUDY DESIGN AND SETTING
The study was a sequential mixed methods research design where data collected in one phase informed the design and focus of the next. Data collection took place between July 2008 and September 2010 at James Cook University, Australia. There were two sections to the study: collection and synthesis of secondary data; and planning, collection and analysis of primary data.

The study began with an exploration of the divide between qualitative and quantitative research. This showed that the divide is more an historical distinction than a current one. As such, there are no theoretical impediments for a single
qualitative and quantitative research CAT. The scope of research methods was examined next through the use of mind maps. This exploration was required so that the design of a CAT could be situated within an overall understanding of research methods. A critical review of how CATs are designed was the final part of secondary data analysis. This review of 45 papers informed the design of the proposed critical appraisal tool, which was based on empirical evidence and the nature of research methods rather than subjective or biased assessments of what a critical appraisal tool could include.

The first part of the primary data collection was an exploratory study of the validity of the scores obtained by the proposed CAT. A random selection of 60 health research papers were analysed using the proposed CAT and five alternative CATs. Next was an exploratory study of reliability, where the proposed CAT was used by five raters, each of whom appraised 24 randomly selected research papers. The final part was to test whether using a CAT was an improvement over using no CAT to appraise research papers because there is little empirical evidence to show if this is true. A total of ten raters were randomly assigned to two groups and they appraised a random selection of five health research papers. One group used the proposed CAT, while the other group did not use any CAT.

**RESULTS**

*Critical review* – Explanations on how a critical appraisal tool was designed and guidelines on how to use the CAT were available in five (11%) out of 45 papers evaluated. Thirty-eight CATs (84%) reported little or no validity evaluation and 33 CATs (73%) had no reliability testing. The questions and statements which made up each CAT were coded into a proposed CAT with eight categories, 22 items, and 98 item descriptors, such that each category and item was distinct from every other.
Validity – In all research designs, the proposed CAT had significant (p < 0.05, 2-tailed) weak to moderate positive Kendall’s tau correlations with the alternative CATs (0.33 ≤ τ ≤ 0.55), except in the Preamble category. There were significant moderate to strong positive correlations in true experimental (0.68 ≤ τ ≤ 0.70); quasi-experimental (0.70 ≤ τ ≤ 1.00); descriptive, exploratory or observational (0.72 ≤ τ ≤ 1.00); qualitative (0.74 ≤ τ ≤ 0.81); and systematic review (0.62 ≤ τ ≤ 0.82) research designs. There were no significant correlations in single system research designs.

Reliability – The intraclass correlation coefficient (ICC) for all research papers was 0.83 for consistency and 0.74 for absolute agreement using the proposed CAT. The G study showed a majority paper effect (53–70%) for each research design, with small to moderate rater effects or paper × rater interaction effects (0–27%).

Compare CAT with no CAT – The ICC for absolute agreement was 0.76 for the group not using a CAT and 0.88 for the proposed CAT group. A G study showed that the group not using a CAT had a total score variance of 24% attributable to either the rater or paper × rater interactions, whereas in the proposed CAT group this variance was 12%. Analysis of covariance (ANCOVA) showed that there were significant effects in the group not using a CAT for subject matter knowledge (F(1,18) = 7.03, p < 0.05 1-tailed, partial η² = 0.28) and rater (F(4,18) = 4.57, p < 0.05 1-tailed, partial η² = 0.50).

DISCUSSION

Critical review – Many CATs have been developed based on a subjective view of research quality rather than on evidence for what elements should or should not be included in a critical appraisal of research. When choosing a CAT, researchers should: (1) take into account the context of the appraisal; (2) determine whether the
CAT was developed using the best evidence available; (3) ensure that the validity of the scores obtained from the CAT can be verified; and (4) analyse the scores obtained from the CAT for reliability.

Validity – The proposed CAT exhibited a good degree of validity based on the theory the CAT was built, the collection of empirical evidence, and the stated context for its use. Therefore, inferences made based on the scores obtained using the proposed CAT should reflect the value of the papers appraised.

Reliability – Given the assessment of validity and the reliability scores obtained, the proposed CAT appears to be a viable tool that can be used across a wide range of research designs and appraisal situations. Any variability in the scores obtained using the proposed CAT can be explained by the diverse subject matter of papers and participants’ unfamiliarity with some research designs. Difficulties with subject matter and research designs are less likely in normal use of the proposed CAT where raters are more familiar with the subject matter and research designs used.

Compare CAT with no CAT – The proposed CAT was more reliable than not using a CAT when appraising research papers. In the group not using a CAT there were significant effects for rater and subject matter knowledge. In the proposed CAT group the rater effect was almost eliminated and there was no subject matter knowledge effect. There was no research design knowledge effect in either group.

CONCLUSION

A CAT was designed and evaluated, which met the aim and objectives of the study. The proposed CAT can be used across a broad range of qualitative and quantitative health research; has the depth to fully assess research papers; has an appropriate scoring system; and has validity and reliability data available. Further research can
extend the proposed CAT to determine whether it is useful in criterion-referencing health research and general research. Furthermore, the proposed CAT can be applied to the increased use of mixed and multiple research methods, and be used to assess, understand and communicate this research knowledge.
Table of contents

Declarations........................................................................................................i
  Release........................................................................................................i
  Copyright ....................................................................................................i
  Ethics ..........................................................................................................i
  Contribution of others .............................................................................. ii
  Incorporated published works................................................................... ii

Acknowledgements ....................................................................................... iv

Abstract .......................................................................................................... v
  Objective ....................................................................................................... v
  Study design and setting............................................................................ v
  Results ......................................................................................................... vi
  Discussion .................................................................................................. vii
  Conclusion ................................................................................................ viii

List of tables .................................................................................................... xv

List of figures .................................................................................................. xvi

Symbols and abbreviations .......................................................................... xvii

Definitions ..................................................................................................... xix

Chapter 1 – Introduction ............................................................................... 1
  1.1 Critical appraisal.................................................................................... 1
  1.2 Critical appraisal tools ........................................................................ 3
  1.3 Aim and objectives ............................................................................... 4
  1.4 Limits to research scope...................................................................... 5
  1.5 Key assumptions .................................................................................. 6
  1.6 Thesis structure ................................................................................... 7
  1.7 References .......................................................................................... 10
# Table of contents

Chapter 2 – Qualitative and quantitative research .......................... 13
  2.1 Abstract .............................................................................................................. 14
  2.2 Introduction ...................................................................................................... 14
  2.3 Research methodology ................................................................................... 15
  2.4 Context, values, and involvement .................................................................. 18
  2.5 Data, analysis, and participants ...................................................................... 20
  2.6 A common error .............................................................................................. 20
  2.7 Conclusion ........................................................................................................ 23
  2.8 In summary ....................................................................................................... 24
  2.9 References ....................................................................................................... 25

Chapter 3 – Research methods .............................................................. 28
  3.1 Abstract ............................................................................................................ 29
  3.2 Introduction ...................................................................................................... 29
  3.3 Research problem ........................................................................................... 32
  3.4 Research design .............................................................................................. 32
  3.5 Sampling technique ....................................................................................... 35
  3.6 Ethical matters ............................................................................................... 37
  3.7 Data collection ............................................................................................... 38
  3.8 Data analysis .................................................................................................. 40
  3.9 Report findings .............................................................................................. 41
  3.10 Conclusion ...................................................................................................... 42
  3.11 In summary ................................................................................................... 43
  3.12 References ..................................................................................................... 44

Chapter 4 – Review of critical appraisal tool design .......................... 46
  4.1 Abstract ............................................................................................................ 47
  4.2 Background ..................................................................................................... 48
  4.3 Methods .......................................................................................................... 51
    4.3.1 Inclusion criteria ..................................................................................... 51
    4.3.2 Exclusion criteria ................................................................................... 52
    4.3.3 Search strategy ...................................................................................... 52
    4.3.4 Ethical matters ....................................................................................... 54
  4.4 Results ............................................................................................................. 54
    4.4.1 Quantitative analysis ............................................................................. 55
    4.4.2 Qualitative analysis ................................................................................. 59
  4.5 Discussion ........................................................................................................ 65
  4.6 Conclusion ....................................................................................................... 68
  4.7 In summary ..................................................................................................... 68
  4.8 References ...................................................................................................... 70
  4.9 Additional material – Search strategy ............................................................ 78
Chapter 5 – Evaluation of validity ................................................. 81
  5.1 Abstract ........................................................................................ 83
  5.2 Introduction .................................................................................. 84
    5.2.1 Construct validity ................................................................. 86
    5.2.2 Validity evaluation ............................................................... 87
      Test content .................................................................................. 88
      Internal structure .......................................................................... 89
      Response processes ....................................................................... 90
      Relations to other variables ....................................................... 91
      Consequences of testing ......................................................... 91
    5.2.3 Study outline ..................................................................... 92
  5.3 Methods ....................................................................................... 92
    5.3.1 Scoring system and user guide .......................................... 93
    5.3.2 Research design .................................................................. 94
    5.3.3 Sample of papers .............................................................. 95
    5.3.4 Data collection and analysis ............................................. 98
    5.3.5 Ethics ................................................................................ 99
  5.4 Results ....................................................................................... 99
    5.4.1 Pre-testing ..........................................................................100
    5.4.2 Main study ..........................................................................100
  5.5 Discussion ................................................................................ 105
    5.5.1 Test content ..................................................................... 105
    5.5.2 Internal structure ............................................................... 106
    5.5.3 Response process ............................................................. 108
    5.5.4 Relations to other variables ............................................. 109
    5.5.5 Consequences of testing ................................................. 109
    5.5.6 Limitations ...................................................................... 110
  5.6 Conclusion ................................................................................ 111
  5.7 In summary ............................................................................... 112
  5.8 References ................................................................................ 113
  5.9 Additional material .................................................................. 117
    5.9.1 User guide for the proposed CAT (evaluation of validity) 117
    5.9.2 Alternative critical appraisal tools .................................. 128
    5.9.3 Worksheet function and decision table ........................... 133
    5.9.4 List of papers used for evaluation of validity ................. 134

Chapter 6 – Reliability study .................................................. 141
  6.1 Abstract .................................................................................... 145
  6.2 Background ................................................................................ 146
  6.3 Methods .................................................................................... 150
    6.3.1 Design ............................................................................... 150
    6.3.2 Data collection .................................................................. 152
  6.4 Results ...................................................................................... 153
    6.4.1 Intraclass correlation coefficient (ICC) ............................. 155
    6.4.2 G and D study ................................................................. 157
    6.4.3 Participant reactions ....................................................... 160
6.5 Discussion ............................................................................................. 161
6.6 Conclusion ............................................................................................. 164
6.7 In summary, .......................................................................................... 165
6.8 References ............................................................................................ 166
6.9 Additional material ............................................................................... 170
  6.9.1 User guide for the proposed CAT (reliability study) ..................... 170
  6.9.2 List of papers used for testing reliability ...................................... 173
  6.9.3. Worksheet function and decision table .................................. 176

Chapter 7 – Compare CAT with no CAT .............................................. 177
  7.1 Abstract ................................................................................................. 178
  7.2 Background ........................................................................................... 179
  7.3 Methods ................................................................................................ 181
    7.3.1 Design ....................................................................................... 182
    7.3.2 Sampling .................................................................................. 182
    7.3.3 Data collection ......................................................................... 184
    7.3.4 Data analysis ........................................................................... 185
    7.3.5 Ethics ........................................................................................ 185
  7.4 Results ................................................................................................... 185
  7.5 Discussion ............................................................................................. 189
  7.6 Conclusion ............................................................................................ 190
  7.7 In summary, .......................................................................................... 191
  7.8 References ............................................................................................ 192
  7.9 Additional material ............................................................................... 196
    7.9.1 Worksheet function and decision table .................................... 196
    7.9.2 Appraisal materials for IA group ............................................ 197
    7.9.3 Appraisal materials for PCAT group ...................................... 199

Chapter 8 – Conclusion ............................................................. 212
  8.1 Design ................................................................................................... 213
  8.2 Evaluation ............................................................................................. 214
  8.3 Limitations ........................................................................................... 218
  8.4 Future research .................................................................................. 219
  8.5 Conclusion ........................................................................................... 220
  8.6 References ............................................................................................ 222

Appendix A – Copyright permissions .................................................. 225
  A.1 Nova Southeastern University.............................................................. 225
  A.2 Springer ............................................................................................... 226
  A.3 Elsevier ................................................................................................. 227
  A.4 Wiley-Blackwell ................................................................................... 229

Appendix B – Ethics approval ................................................................. 230
Appendix C – Published articles .................................................. 232
   C.1 Qualitative and quantitative research designs are more similar than different (Chapter 2) .................................................. 233
   C.2 Mind mapping research methods (Chapter 3) ................................................. 241
   C.3 A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed (Chapter 4) ........................................... 253
   C.4 A general critical appraisal tool: An evaluation of construct validity (Chapter 5) .............................................................................. 264
   C.5 Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs (Chapter 6) ................... 276
   C.6 Comparison of the effects of using the Crowe Critical Appraisal Tool versus informal appraisal in assessing health research: A randomised trial .......................................................... 285

Appendix D – Material for participants, reliability study ................. 291
   D.1 Informed consent form .................................................................. 292
   D.2 Information sheet ........................................................................ 293
   D.3 Questions for participants ............................................................... 294

Appendix E – Material for participants, compare CAT with no CAT ... 295
   E.1 Informed consent form .................................................................. 296
   E.2 Information sheet ........................................................................ 297
   E.3 Pre-appraisal questions ................................................................ 298
   E.4 Post-appraisal questions ................................................................. 299

Appendix F – Crowe Critical Appraisal Tool .................................... 300
   F.1 Crowe Critical Appraisal Tool (CCAT) ........................................... 301
   F.2 Crowe Critical Appraisal Tool (CCAT) user guide ......................... 303

Complete reference list ............................................................. 315
   Works cited ........................................................................................... 315
   Papers appraised .................................................................................. 327

Full mind map of research methods (see Chapter 3) .... Last page insert
List of tables

Table 2.1 Hierarchy of evidence for quantitative studies ........................................ 22
Table 4.1 Search terms and databases ................................................................ 53
Table 4.2 Summary of critical appraisal tools .................................................... 56
Table 4.3 Categories and items included in CATs ............................................. 62
Table 5.1 Paper search strategy ...................................................................... 97
Table 5.2 Proposed CAT structure after initial pilot ........................................ 101
Table 5.3 Average scores for proposed CAT vs alternative CATs .................. 102
Table 5.4 Kendall’s tau for proposed CAT vs alternative CATs ..................... 103
Table 6.1 Proposed critical appraisal tool (CAT) .......................................... 148
Table 6.2 Summary of ICCs (k = 4, excludes Rater III) .................................. 156
Table 6.3 Percentage mean variance components (k = 4, excludes Rater III) ... 158
Table 6.4 D study (excludes Rater III) .......................................................... 160
Table 7.1 Reliability (total score %, k=5, n=5) ............................................... 188
Table 7.2 Analysis of covariance ................................................................... 188
List of figures

Figure 2.1 Inductive and deductive reasoning .......................................................18
Figure 2.2 Research designs ............................................................................. 21

Figure 3.1 Research methods ............................................................................ 31
Figure 3.2 Research problem ............................................................................. 32
Figure 3.3 Research design ............................................................................... 33
Figure 3.4 Sampling technique ......................................................................... 36
Figure 3.5 Ethical matters ................................................................................. 37
Figure 3.6 Data collection ............................................................................... 39
Figure 3.7 Data analysis ................................................................................. 41
Figure 3.8 Report findings .............................................................................. 42

Figure 4.1 Flow-diagram of search results .......................................................54

Figure 7.1 Flow of participants ..................................................................... 186

Full mind map of research methods (see Chapter 3).................................. Last page insert
Symbols and abbreviations

$E\rho^2$ – Relative generalizability coefficient

Also: Relative G coefficient; Generalizability coefficient; G coefficient

Alternative: $E\rho_0^2$; $\rho_0^2$.

$\Phi$ – Absolute generalizability coefficient

Also: Absolute G coefficient; Index of dependability

Alternative: $\rho_\Phi^2$

AERA – American Educational Research Association

AMSTAR – Assessment of Multiple Systematic Reviews

APA – American Psychology Association

CASP – Critical Appraisal Skills Programme

CAT – Critical appraisal tool

CCAT – Crowe Critical Appraisal Tool

CEBM – Centre for Evidence-Based Medicine

CHE – Centre for Health Evidence

CONSORT – Consolidated Standards for Reporting of Trials

COREQ – Consolidated Criteria for Reporting Qualitative Research

CRD – Centre for Reviews and Dissemination

CTT – Classical Test Theory

D study – Decision study

DEO – Descriptive, exploratory or observational research designs

DOI – Digital object identifier
EBP – Evidence-based practice
EMS – Expected mean square
G coefficient – Generalizability coefficient
G study – Generalizability study
G theory – Generalizability theory
IA – Informal appraisal
ICC – Intraclass correlation coefficient
IRT – Item response theory
JCU – James Cook University
MOOSE – Meta-analysis of Observational Studies in Epidemiology
NCME – National Council on Measurement in Education
NCMUE – (OBSOLETE, SEE NCME) National Council on Measurements Used in Education
NHMRC – National Health and Medical Research Council
PCAT – Proposed critical appraisal tool
PEDro – Physiotherapy Evidence Database
PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-analyses
QUOROM – (OBSOLETE, SEE PRISMA) Quality of Reporting of Meta-analyses
RCT – Randomised controlled trial
RSS – Really Simple Syndication
SQUIRE – Standards for Quality Improvement Reporting Excellence
STROBE – Strengthening the Reporting of Observational Studies in Epidemiology
Definitions

Critical appraisal
The impartial assessment of one or more research papers to determine their strengths, weaknesses, and benefits. Where,

1. Strengths – Suitability of research methods to answer the research question.
2. Weaknesses – Identification and, where possible, reduction of limitations due to research methods.
3. Benefits – Implications based on sound conclusions drawn from the research methods used, results obtained, and current evidence.

Critical appraisal tool
A structured approach to critical appraisal.

Research design
The basic approach or approaches used to answer a research question, such as true experimental or phenomenological designs. Research design is one element of research methods.

Research methodology
The philosophical (ontological) and theoretical (epistemological) basis for research designs.
**Research methods**

The overall process of initiating, implementing, analysing, and reporting research. The term is always used in the plural. Elements of research methods are research question, research design, sampling techniques, ethical matters, data collection, data analysis, and report findings.

**RSS (Really Simple Syndication)**

A standardised method to periodically and automatically download frequently updated information from a source connected to the internet. Also known as a feed, web feed, or channel.
Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality.

*Edward Thorndike (1918)*