1. Critical appraisal: how to examine and evaluate the research evidence

Alyson L. Dodd, Siobhan Reilly, Faraz Ahmed and Carol Thomas

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context. (Burls 2009, p. 1)

INTRODUCTION

This chapter on critical appraisal provides an essential guide to students and researchers on the what, why and how of an increasingly important set of skills in health research. Indeed, critical appraisal is now a key tenet of academic writing and scholarship. In large measure, the quality of academic writing is judged by how authors use evidence to form arguments and demonstrate critical thinking (Moon 2005). As well as writing, this important aspect of scholarship also includes reading and understanding (Evans 2016). Throughout their studies, students will often be asked to critically evaluate, critically review or critique published materials; these are all different ways of saying that critical appraisal is required.

This chapter covers:

- what critical appraisal is;
- why critical appraisal is important;
- critical appraisal tools and checklists, including an overview of how to find and select them for different research designs; and
- useful practical guidance on how to assess methodological quality – what, why, when and how.

LEARNING OUTCOMES

After studying this chapter, you should be able to:

- understand what critical appraisal is, particularly in relation to empirical health research studies;
- appreciate why and when critical appraisal is relevant;
- identify and find critical appraisal tools for different types of research designs; and appreciate that not all research outputs should be given equal weight.

WHAT IS CRITICAL APPRAISAL?

Critical appraisal focuses primarily on the methodological quality of empirical work,¹ and its relevance to a researcher's context (see also Chapter 2 in this volume, on evidence synthesis). This applies whether the context is a university course, a research project or clinical practice (Keen and Otter 2014). This quality assessment involves asking how appropriate a study design is for exploring a specified research question, and the rigour of the approach used. Specifically, it is about the validity, reliability and utility of research (see Box 1.1 for an example of how these principles are applied in critical appraisal, and later chapters for more details of validity and reliability). Another important aspect of critical appraisal is establishing relevance, and assessing whether findings are really saying what the authors claim they are saying. Are the conclusions that researchers draw a fair representation of the work undertaken (from design, to data collection, to data analysis)? This links back to a study's initial research questions, aims and objectives and, ultimately, to the theoretical framework employed. So, while critical appraisal focuses on methodological rigour, the importance of the role of theory in research should not be

BOX 1.1 ASSESSING VALIDITY, RELIABILITY, UTILITY AND RELEVANCE

A starting point for understanding critical appraisal is familiarity with the meaning of a set of concepts about the quality of published research studies: validity, reliability, utility, and relevance. The example below is designed to introduce these concepts, and to pose some questions you should ask yourself to ensure these are included. (Note: publication of research does not necessarily mean that a research study has all of these attributes.)

Imagine you are conducting a review that aims to make an evidence-based argument about the hypothesis that rumination (that is, repetitive negative thinking) is positively associated with depression. Here are some key questions to ask yourself initially in order to appraise a paper identified in a literature search.

1. Validity questions:
 - (a) Does the study assess rumination and depression using suitable validated measures? (Test validity)
 - (b) Are the measures used observer-rated or self-report?
 - (c) Are potentially confounding variables controlled for? (That is, does the study minimize bias?) (Internal validity)
 - (d) Does the study recruit individuals with a clinical diagnosis of depression, or measure depressive symptomatology in a non-clinical (analogue) population? (External validity/generalizability)
2. Reliability questions:
 - (a) Are the measures used themselves reliable? (Internal consistency)
 - (b) Is the same result likely to be found again? (Test-retest reliability)
 - (c) If an observer-rated measure of depressive symptoms is used, is this consistent for the same participant across different raters? (Inter-rater reliability)
3. Utility question:
 - (a) Does the study have practical significance for your review?
4. Relevance questions:
 - (a) Is the research question relevant to the focus of your review?
 - (b) Is the population of interest appropriate for your review?
 - (c) Are the conclusions related to the hypothesis?

underestimated (although the role of theory differs when considering quantitative and qualitative research; see Parts II and III in this volume).

WHY DO RESEARCHERS NEED TO UNDERTAKE CRITICAL APPRAISAL?

Beyond research and academia, critical appraisal is a valuable life skill; we all need to keep an air of criticality about us in all aspects of life. For example, if you want to learn about a new topic, it would not serve you well to read just one book and treat that as ‘the truth’; instead, you would read around and examine the evidence presented across multiple resources and perspectives.

Critical appraisal helps us to identify bias and assess whether we believe authors’ claims (Keen and Otter 2014). We use critical appraisal when we want to know what the evidence is really saying. Poor methodological rigour and skewed conclusions in evidence are common concerns in everyday life. However, health research of the highest quality does not definitively prove or disprove a theory or hypothesis; it can only support or refute it.

Unfortunately, mass media reporting often mistakenly infers causality, and sometimes researchers can fall into the same trap. Knowledge must be presented appropriately, not just for academic audiences but for everybody (Evans 2016). Research that is poorly conducted and/or poorly reported is sometimes unquestioningly accepted by the general public, and can end up being implemented in health services. A recent example is Study 329, a 2001 clinical trial of the anti-depressant paroxetine in adolescents (Box 1.2).

Greenhalgh (2019, p.29) proposes that as many as 99 per cent of published papers are of poor quality, stating: ‘bad science is bad science regardless of whether the study addressed an important clinical issue, whether the results are “statistically significant”, whether things changed in the direction you would have liked them to, and whether, if true, the findings promise immeasurable benefits for patients or savings for the health service’. This emphasizes the importance of critical appraisal. As a researcher, you are not looking for evidence that supports what you already believe to be true. Instead, you are

BOX 1.2 EXAMPLE OF POOR METHODOLOGY AND REPORTING: STUDY 329

The Restoring Invisible and Abandoned Trials (RIAT) initiative (Doshi et al. 2013) called for researchers to publish results from clinical trials that had previously not been published, or to correct findings that had been published but where findings were biased.

Study 329 was a randomized controlled trial testing the effectiveness of paroxetine (an anti-depressant commonly known as Seroxat in the UK) for the treatment of depression in adolescents (Keller et al. 2001). The trial had a history of controversy and the data were re-analysed as part of RIAT. In the re-analysis (Le Noury et al. 2015), the authors reported that, not only was paroxetine not superior to placebo, but it actually increased rates of suicidal thoughts and behaviours.

This is an example emphasizing that even research published in peer-reviewed medical journals should not be taken at face value, as there may be selective reporting of findings.

looking for evidence that appraises each paper for its merits. To be a successful researcher, you must be able to assess published papers with a critical eye, and apply this critical eye to your own research at all stages. This approach will help you when conducting a critical evaluation of the literature (moving towards a systematic review and evidence synthesis; see Chapter 2).

When you come to design your own research, critical appraisal across multiple studies relevant to your topic allows you to methodically identify the gaps in the literature. Empirical work found in a systematic search of the relevant literature should be appraised individually and in relation to one another, in order to synthesize findings and give a balanced systematic (or narrative) review of the evidence (see Chapter 2). For example, it is important to identify heterogeneity across studies, by identifying whether they have used the same measures of the same constructs, have recruited from the same population, have conducted the same statistical tests or have a similar control group (if conducting an experiment or randomized controlled trial: see Chapter 12). Some studies may have used a more robust approach than others, and this would be uncovered during critical appraisal. It can be hard to compare like for like (see Gagnier et al. 2012 for further discussion on clinical heterogeneity).

Sometimes a systematic literature review on the same or a similar topic will already have been published. Scoping searches will help you to identify these and refine your review question (see Chapter 2). As with any published research, these reviews should not be ignored and should be appraised for their methodological rigour. You need to ask, how can I make sure my research is saying something new or updating these findings?

Critical appraisal of existing empirical work on your topic, or your methodology, will help you to develop your research questions (see Chapter 3) and design your research with high-quality methods to ensure the generalizability, reliability and validity of your study. Being able to demonstrate the methodological rigour of your research and its contribution to the literature is particularly important if you come to write an application for funding from a research grant. When you are convincing the funding panel to award you a significant sum of money, they need to be assured that there is evidence to support the importance of your research and, equally, that you are not repeating unnecessary research. This is not only relevant to funding applications; your original contribution to knowledge should be clear regardless of the type of research, that is, whether it is for your PhD, a small pilot study or a large-scale externally-funded project or programme of research.

Critical appraisal is also vital when writing up your findings and peer-reviewing other people's findings for publication in an academic journal. In this sense, critical appraisal comes full circle.

Summary

Critical appraisal enables researchers to:

- Stop making unfounded assumptions about the quality of a research project. Critical appraisal involves 'not knowing all the answers before you start' (Greenhalgh 2019, p.11).
- Identify the strengths, weaknesses, biases and heterogeneity in research in a systematized way.

- Undertake an appraisal of appraisals, that is, existing reviews on the same or a similar topic should not be ignored.
- Justify your research and be clear where it sits within the literature from proposal to dissemination of findings.
- Design your research. Ask yourself: what has already been done? Was it methodologically rigorous? Do the conclusions match the findings? What requires replication?
- Write (and review) journal articles and grant applications objectively.
- Relate already published research to their local situation and/or research with precision. For example, if their review of evidence is on the psychological mechanisms underlying health concerns, then relevant types of evidence are not those that explore underlying biological mechanisms or service user experiences.
- Facilitate the implementation of effective interventions in their practice or policy situation through an understanding of the current evidence-base.

OVERVIEW OF CRITICAL APPRAISAL TOOLS AND CHECKLISTS: HOW TO FIND AND SELECT THEM FOR DIFFERENT RESEARCH DESIGNS

This section is designed to assist you to become a critical consumer of research, in preparation for planning research of the highest quality. As noted in Chapter 2, until recently, when writing a literature review, researchers would traditionally have gathered pertinent information about a topic in a largely unstructured way, potentially omitting any critique of papers, and cherry picking those papers that supported the pre-held views of the researcher. The evidence-based medicine movement has, in recent years, encouraged those undertaking a literature review to utilize more systematic approaches, although not all literature reviews qualify as a systematic literature review (see Chapter 2 for a discussion of different approaches to reviewing the literature). The quantity of reading that a literature search turns up can be daunting and frustrating. Indeed, having hundreds of papers waiting to be read, and made sense of, is a much more difficult process without the use of critical appraisal tools and skills. In the same way as regulatory organizations (such as the Care Quality Commission in England) will have various discriminating questions they need answering, researchers too need to be armed with some questions. Without these you will be like a joiner without a screwdriver, a writer without a pen, and so on. The tools or checklists of criteria enable you to critically assess the quality of a published research study in a more structured, rigorous, standardized and minimally biased manner.

Recently there have been many developments, and the choice of appropriate appraisal tools has become crucial. These tools facilitate critical appraisal of different types of empirical research, generally using a set of criteria or a checklist allowing for a score or quality rating to be derived, using rating scales or tick boxes. The tool you select and the depth used in appraising studies will depend on the type of review and its aim. If you are conducting a systematic review, this generally involves an in-depth quality assessment of included studies. If you are conducting a mapping or scoping review however, this would map out and categorize existing literature and identify gaps in the literature (Grant and Booth 2009), but not necessarily require a critical appraisal of each element of that body of

literature. If you are conducting a literature review, then you will need to critically appraise some or all of the studies that you include in your review (see Chapter 2). If you are adopting a more systematic type of approach then you will need to specify your procedures in detail. If you want to assess the quality of studies for a literature review adopting a systematic approach, then you need to make sure you choose an appropriate tool. However, you might just want to use the checklists or tools to highlight aspects of the studies you want to write about but not actually analyse them in a structured or systematic way.

As there are many different tools to choose from some thought is required before you make your selection. The tool needs to be suitable for your review or prospective research design. There is, therefore, a need to have some understanding of different research designs beforehand. Critical appraisal tools help you to formulate your views and opinions on a piece of research you read about in a structured, more objective, manner. The tools can also enable data tables to be constructed; assisting with the structuring and organization of how the assessments are written up.

The tool you select also depends on what types of studies you are reviewing. If the review focuses upon a single method, the choice of tool will be easier. Things get increasingly difficult when multiple methods are included, and tools designed to tackle multiple-method reviews may feel over-complicated, unwieldy and less satisfactory. Therefore, it can be preferable to use a critical appraisal tool for each research design instead of an all-inclusive tool.

Some tools use a scoring method for quality, making them less attractive than to those using a checklist, as these can provide more useful information about quality of a particular study. A numerical value does not provide any detail about the individual elements of the quality assessment. As with any method, it is also important to acknowledge the limitations in any critical appraisal tool or assessment of methodological quality.

HOW TO DO CRITICAL APPRAISAL? IDEAS FOR SOURCING APPROPRIATE TOOLS AND REPORTING GUIDELINES

- Examine other reviews in similar topic areas, to see what others have used.
- Look at the resources suggested in Table 1.1; these are not exhaustive lists, but will get you started.
- Look at the recommended reporting guidelines for individual study types, for example, see the Equator network website² (see Table 1.1 for further details).
- See examples of systematic review guides, such as www.liv.ac.uk/systematic-review-guide.
- It is also worth remembering that some types of studies will have clear reporting guidelines (see the Equator network website). You should explore these for the types of studies you are interested in.
- See web resources in Boxes 1.3 and 1.4.

WHICH TOOL TO USE AND HOW TO DECIDE?

The literature on critical appraisal and tools is an evolving and dynamic field. This requires reviewers to be up to date with the latest thinking. Even the tools are not static

Table 1.1 Reporting guidelines for main study types

Type of study	Guideline	
Randomized trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Case reports	CARE	
Qualitative research	SRQR	COREQ
Diagnostic/prognostic studies	STARD	TRIPOD
Quality improvement studies	SQUIRE	
Economic evaluations	CHEERS	
Animal pre-clinical studies	ARRIVE	
Study protocols	SPIRIT	PRISMA-P

Source: <http://www.equator-network.org/> (accessed 4 March 2020).

documents and will continue to go through development during the course of their life cycle. As a reviewer, you will have to explore whether there are other alternates or an updated iteration of tools that have better congruence with your review aim, review question and data available. So which tool should you use and how to decide? For the first part of this question you will need to know which relevant tools are available and where to find them, and have an idea about how to use them. For the second part of the question you will need to know what type of study design(s) you are expecting to review, how many studies you have to review and how much time you have available to review them. There are currently a wide variety of tools available, too many to list here. Unfortunately there is not yet a single methodological quality tool that is suitable for use in all reviews of health research. In Table 1.2 we have provided basic details on a number of tools and what types of designs they cover. We recommend that the reader explores tools for their specific needs, recognizing that tools are being developed and updated regularly.

Even though there are a variety of tools, many researchers still opt to use the CASP tools (Critical Appraisal Skills Programme 2020). The CASP may suit the purpose exactly, but could also be considered a lazy option given that this tool appears first on a Google search of ‘critical appraisal tools’. However, if you want to assess a study for clinical practice purposes, your criteria for selecting a critical appraisal tool will be different, and the CASP tools could be a sensible choice.

Where do you find other tools? There are a number of approaches you can take. For example, you could identify which tool the authors of a previous systematic review in your topic area used, but be aware that this does not always mean that a good decision was made. You also need to be sure that the tool works for your review question, task or situation.

There have been a number of reviews on this subject that are particularly useful, as they highlight the differences between tools (for example, Katrak et al. 2004). These authors note that ‘the selected tools should have published evidence of the empirical basis for their construction, validity of items and reliability of interpretation, as well as guidelines for use, so that the tools can be applied and interpreted in a standardized manner’ (Katrak et al. 2004, p.8). They also noted that the most common items included in the critical appraisal tools that they appraised were:

Table 1.2 A selection of critical appraisal tools and checklists

Name, reference	Authors, source	Content, areas/domains assessed	Length	Comments (limitations and presence of guides)
Qualitative tools/checklists				
McMaster Critical Review Form – Qualitative Studies (Version 2.0)	Letts et al. (2007)	Includes sections on design, theoretical/philosophical perspectives and methods	12 sections covering common research study headings; 21 questions	Simple layout; easy to understand questions; completion guidance included in the comments boxes to aid completion Separate guidance document: http://srs-mcmaster.ca/wp-content/uploads/2015/04/Guidelines-for-Critical-Review-Form-Qualitative-Studies-English.pdf For use only with interviews (in-depth or semi-structured) and focus groups as these methods are the most common methods for data collection in qualitative health research Easy to follow instructions/questions Long at 32 questions Aims to consolidate reporting of qualitative studies
COREQ: Consolidated Criteria for Reporting Qualitative Studies	Tong et al. (2007)	The checklist is based on three domains: 1. Research team and reflexivity 2. Study design 3. Analysis and findings	A 32-item checklist, can help researchers to report important aspects of the research team, study methods, context of the study, findings, analysis and interpretations. The 32 items are based on 76 items from 22 checklists to assess the integrity of qualitative studies covering in-depth interviews and focus groups	

Specialist Unit for Review Evidence (SURE)	Specialist Unit for Review Evidence (SURE) (2013)	Includes sections on design, research team, methods, ethical consideration and data analysis	Ten key question areas with many subsidiary questions	For use with all forms of qualitative study No additional guidance but questions appear easy to understand Limited explanation on its development except that it was adapted and updated from the former Health Evidence Bulletins Wales (HEBW) checklist with reference to the NICE Public Health Methods Manual (2012) and previous versions of the Critical Appraisal Skills Programme (CASP) checklists
Standards for Reporting Qualitative Research: A Synthesis of Recommendations (SRQR)	O'Brien et al. (2014)	Checklist assesses six main areas: 1. Title and abstract 2. Introduction 3. Methods 4. Results/findings 5. Discussion 6. Other information	A checklist of 21 standards to help guide researchers on how to report qualitative studies. Collated from 40 separate sources. Each standard in the checklist is essential for complete, transparent reporting of qualitative research	For use with all qualitative research designs. Easy to understand questions. Supplemental guidance is found in its appendix 2, available as extra digital content at http://links.lww.com/ACADMED/A218 , which contains a detailed explanation of each standard, together with examples from recently published qualitative studies (Guidance advises that it is inappropriate to use the SRQR to judge the quality of research methods and findings as standards do not define methodological rigour)
Enhancing Transparency in Reporting the Synthesis of Qualitative Research: ENTREQ	Tong et al. (2012)	Checklist contains five main domains: 1. Introduction 2. Methods and methodology 3. Literature search and selection 4. Appraisal 5. Synthesis of findings	Checklist contains 21 items, developed inductively from guides to synthesizing qualitative health research, seminal methodology papers and the authors' experience in conducting and appraising qualitative syntheses	For use when reporting the synthesis of qualitative studies; checklist was developed to promote explicit and comprehensive reporting of the synthesis of qualitative studies Easy to understand as guidance is included

Table 1.2 (continued)

Name, reference	Authors, source	Content, areas/domains assessed	Length	Comments (limitations and presence of guides)
Appraising Qualitative Research in Health Education: Guidelines for Public Health Educators	Jeanfreau and Jack (2010)	Guidance explores seven main areas: 1. Publication title, 2. Abstract, 3. Introduction, 4. Methods, 5. Data analysis, 6. Results, 7. Discussion/conclusions	Guidance suggests six questions that may guide the evaluation of qualitative research articles in order to determine the strengths and weaknesses of the methodology, to evaluate the quality or strength of the study's evidence, and to identify the appropriateness for use in the reader's practice	Provides an overview of qualitative research approaches, defines key terminology used in qualitative research, and provides guidelines for appraising the strengths and weaknesses of published qualitative research For use with all forms of qualitative study Easy to understand as guidance included
NICE: Methodology Checklist for Qualitative Studies	National Institute for Health and Care Excellence (NICE) (2012)	Checklist contains six main domains: 1. Theoretical approach 2. Study design 3. Data collection 4. Validity 5. Analysis 6. Ethics	Checklist is based on three key checklists: 1. Quality in qualitative evaluation: a framework for assessing research evidence 2. Critical Appraisal Skills Programme (CASP) 3. National Training and Research Appraisal Group (NTRAG) checklist	For use with all forms of qualitative study Easy to understand as guidance is included, and easy to follow instructions/questions. Concise, includes 12 sub-questions items (with six main themes for scorings)

Quantitative tools/checklists			
EPHPP Quality Assessment Tool for Quantitative Studies	Thomas et al. (2004)	<p>Selection bias</p> <p>Study design</p> <p>Confounders</p> <p>Blinding</p> <p>Data collection methods</p> <p>Withdrawals and dropouts</p> <p>Analyses</p> <p>Internal validity (research question, literature search, listing of included/excluded study, quality of included study, methods, bias)</p> <p>Assessment of methodological quality</p>	<p>Covers many designs; although the tool is designed for quantitative studies, it seems to be focused on effectiveness studies</p> <p>Randomized controlled trial</p> <p>Controlled clinical trial</p> <p>Cohort analytic (two group pre + post)</p> <p>Case control</p> <p>Cohort (one group pre + post)</p> <p>Interrupted time series</p> <p>Other specify</p> <p>Tool was developed for use in public health, and can be applied to articles of any public health topic area, including the promotion of family and sexual health and the prevention of chronic disease, injuries and substance misuse. Various types of public health professionals would find this tool relevant to utilize sources of high-quality literature to support the decision-making process, especially when designing, implementing and evaluating public health programmes and policy</p> <p>There is also a 'Quality assessment tool for quantitative studies dictionary' to describe items in the tool, thereby assisting raters to score study quality (see https://merst.ca/wp-content/uploads/2018/02/quality-assessment-dictionary_2017.pdf)</p>

Table 1.2 (continued)

Name, reference	Authors, source	Content, areas/domains assessed	Length	Comments (limitations and presence of guides)
Cochrane Risk of Bias Tool	Higgins et al. (2011)	Cochrane Collaboration's tool for assessing risk of bias contains questions on six forms of bias: 1. Selection 2. Performance 3. Detection 4. Attrition 5. Reporting 6. Other	Seven main questions on bias	There is guidance on completing the questions The tool is mainly suitable for use with clinical trials/studies
Evaluation Tool for Quantitative Research Studies	Long et al. (2002b)	Assesses quantitative papers on five key areas: 1. Overview 2. Study 3. Ethics 4. Comparable group 5. Policy and practice implications		No guidance with it Total questions amount to 51, which is long Unclear on its development, reliability or validity

Mixed methods tools/checklists					
A Scoring System for Mixed Methods Research and Mixed Studies Reviews	Pluye et al. (2009, 2011)	Qualitative (six items), quantitative experimental (three items), quantitative observational (three items) and mixed methods (three items)	Reliability assessment was described in two of the 17 systematic mixed studies reviews (SMSRs): http://www.nccmt.ca/resources/search/232	<p><i>Content has been validated using feedback from experts and workshops</i></p> <p>No mention of specific bias assessment although the quality appraisal uses ranking by methodological quality</p> <p>No valid checklists are used by the SMSR to appraise studies</p> <p>A scoring system is proposed for concomitantly appraising the methodological quality of qualitative, quantitative and mixed methods studies for SMSRs. This may also be used to appraise the methodological quality of the various research</p> <p>There is a tutorial on how to use the tool</p> <p>The tool is still in development but is being used widely</p>	
Evaluation Tool for 'Mixed Methods' Study Designs	Long et al. (2002a)	Study overview, context, ethics, group comparability, qualitative data, policy and practice implications and other	Many sub-questions	<p>Critical appraisal tool for qualitative studies. Developed by Long et al. at the Health Care Practice R&D Unit (HCPRDU), University of Salford as part of a feasibility study for examining the feasibility of undertaking systematic reviews in social care</p> <p>No guidance</p> <p>No assessment of validity or reliability of tool</p>	

Note: All websites accessed 3 March 2020.

BOX 1.3 RESOURCES FOR QUALITY ASSESSMENT (QUANTITATIVE)

- Critical Appraisal Skills Programme (CASP) checklists: <http://www.casp-uk.net/>.
- Cochrane Collaboration's tool for assessing risk of bias in randomized trials: Higgins et al. (2011).
- Key reporting guidelines (including CONSORT statement) Enhancing the Quality and Transparency of Health Research (Equator): <http://www.equator-network.org/>.
- AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of health-care interventions, or both: Shea et al. (2017).
- Best Evidence Topics (BETs) Critical Appraisal Worksheets: <http://www.bestbets.org/links/BET-CA-worksheets.php>.
- Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields: Kmet et al. (2004).

Note: All websites were accessed 4 March 2020.

BOX 1.4 RESOURCES FOR QUALITY ASSESSMENT (QUALITATIVE)

- Cochrane Qualitative and Implementation (CQIMG) Supplementary Guidance on Quality Assessment: <http://methods.cochrane.org/qi/supplemental-handbook-guidance>.
- Critically Appraising Qualitative Research (BMJ article): Kuper et al. (2008).
- CASP Qualitative Research Critical Appraisal Checklist (revised edition May 2013): <http://www.casp-uk.net/>.
- JBI Checklist for Qualitative Research (Joanna Briggs Institute, 2017: https://joannabriggs.org/ebp/critical_appraisal_tools).
- McMaster Tool for Qualitative Studies (Letts et al. 2007; version 2).
- Quality in Qualitative Evaluation (Spencer et al. 2003).
- Quality of Reporting Tool (Carroll and Booth 2015).
- Quality Prompts (Dixon-Woods et al. 2004).
- Hawker et al. (2002): Qualitative study assessment guidelines and quality appraisal form.

Note: All websites accessed 4 March 2020.

- Eligibility criteria (inclusion/exclusion criteria) (N = 63)
- Appropriate statistical analyses (N = 47)
- Random allocation of subjects (N = 43)
- Consideration of outcome measures used (N = 43)
- Sample size justification/power calculations (N = 39)
- Study design reported (N = 36)
- Assessor blinding (N = 36).

However, Katrak et al.'s (2004) review only included critical appraisal tools if they: were applicable to at least one research design (quantitative and qualitative research), had clear and unambiguous criteria, and could produce a numeric quality score. Ideally the psychometric properties of tools would also be established; however, intra-rater reliability was established for only one critical appraisal tool (Downs and Black 1998) as part of its empirical development process.

Critical appraisal has become a growth area over the past decade or so, and therefore choosing a suitable tool is now more difficult as there are more tools available to consider. You may also feel that, despite this, the tools available are not appropriate for your review and you may wonder about making your own new checklist. The temptation can be to pick and mix from existing tools, but remember, without validation this will not be a validated tool and you will not be able to compare the findings with other areas. Although many existing tools are not validated, justifying this can be problematic, especially within the PhD viva situation, or to a journal reviewer, especially if the items you chose to omit include important measures of quality for the studies you are examining. We would therefore recommend not making up your own tool.

Features of tools that you may want to consider when making your choice are indicated in Table 1.2.

SUMMARY

This chapter has introduced the what, why and how of critical appraisal, and has underlined the importance to research studies of the concepts validity, reliability, utility and relevance. A key message is that published research should not be taken at face value. The discussion then moved on to an examination of critical appraisal tools and checklists. One of the major advantages of critical appraisal is the assistance it provides when trying to understand previous research, so that study results can be applied appropriately. This ensures that future resources are not wasted by repeating the same mistakes or by repeating the same research.

Patients have suffered and died unnecessarily and resources for research have been wasted *because the research community has failed to review existing evidence systematically* when planning new research. *Why should patients and the public trust us* if we and our professional institutions fail to make systematic, efficient use of the results of the research that the public has funded? (Chalmers 2014, emphasis added)

Another advantage is that critical appraisal skill builds up your methodological knowledge in your research area. Observing the mistakes of others enables researchers to avoid making them in their own research. It also alerts them to the feasibility of what is possible. Examining the work of previous researchers, and noting what resources were available for that study, is a useful check on what you might be proposing in your own study. If a study was funded by a large grant, with a number of researchers working in different sites, it is unlikely that you would be able to conduct a similarly sized study without similar resources.

The advantages of critical appraisal are that it:

- helps with your understanding of a study, providing a structured way of reading and interpreting;
- helps to identify the quality of a study;
- helps you to draw meaningful conclusions about individual studies and overall bodies of research; and
- allows you to develop long-lasting critical appraisal skills that, over time you will apply more naturally to your reading of the research evidence.

BOX 1.5 LEARNING ACTIVITIES: ASSESSING THE METHODOLOGICAL QUALITY OF HEALTH RESEARCH

1. Identify other methodological quality appraisal tools in addition to those in Tables 1.1 and 1.2.
 - (a) Organize these into a format that will help you make an informed decision about the tools. (When could they be used and in what context? For which research designs?)
 - (b) This could be a structured document, a table or a bullet-point list. Areas that could be included: appropriate study designs, number of items/page length, ease of completion, validity/reliability, and so on.
2. Identify two studies within scope of your review (use Chapter 2 on evidence synthesis to help you with reviewing the literature).
 - (a) Identify a tool that is appropriate for the studies selected.
 - (b) Complete this tool for both studies.
 - (c) Think about the reasons why you chose your tool and how you found the process of assessing methodological quality. This may also include any problems or advantages you experienced when using these tools, and any other relevant points of discussion and observation.
3. Find a recent review that you are interested in (consider choosing a Cochrane or a Campbell Collaboration as they tend to include details on individual studies).
 - (a) Choose two papers that have been included in the review.
 - (b) Identify the quality appraisal tool the authors used.
 - (c) Complete the quality appraisal tool and compare your results with those that the authors have made. How did you do?
4. Set up a journal club/group with fellow students, practitioners, and so on.

The disadvantages of critical appraisal are that:

- undertaking the process is often not clean-cut. It can be messy and needs some subjectivity, and a second reviewer if possible;
- it requires lots of notes and recording what you are doing and what you have found; and
- reviewer experience varies and this has an effect on critical appraisal/quality assessment (again, two reviewers can limit this effect).

NOTES

1. The UK organization, Vitae, who focus on nurturing and developing leading researchers, list critical thinking and evaluating as key characteristics of 'effective researchers' in their Researcher Development Framework (2010), highlighting the importance of acquiring these skills (see <https://www.vitae.ac.uk/>).
2. Equator network website: <http://www.equator-network.org/> (accessed 4 March 2020).

REFERENCES

- Burls, A. (2009), 'What is critical appraisal?', *Hayward Medical Communications*, accessed 1 June 2016 at http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_is_critical_appraisal.pdf.
- Carroll, C. and A. Booth (2015), 'Quality assessment of qualitative evidence for systematic review and syn-

- thesis: is it meaningful, and if so, how should it be performed?', *Research Synthesis Methods*, **6** (2), 149–54, doi:10.1002/jrsm.1128.
- Chalmers, I. (2014), 'How to increase value and reduce waste when research priorities are set', paper presented at Reward Alliance symposium, London, 8 January, accessed 7 September 2017 at <http://researchwaste.net/documents/presentations/>.
- Critical Appraisal Skills Programme (CASP) (2020), 'CASP checklists', accessed 4 March 2020 at <https://casp-uk.net/casp-tools-checklists/>.
- Dixon-Woods, M., R.L. Shaw, S. Agarwal and J.A. Smith (2004), 'The problem of appraising qualitative research', *Quality and Safety and Health Care*, **13** (3), 223–5.
- Downs, S. and N. Black (1998), 'The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions', *Journal of Epidemiology and Community Health*, **52** (6), 377–84.
- Doshi, P., K. Dickersin, D. Healy, S.S. Vedula and T. Jefferson (2013), 'Restoring invisible and abandoned trials: a call for people to publish the findings', *BMJ*, **346** (7913), f2865.
- Evans, J. (2016), *How to Be a Researcher: A Strategic Guide for Academic Success*, Hove: Routledge.
- Gagnier, J.J., D. Moher, H. Boon, J. Beyene and C. Bombardier (2012), 'Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature', *BMC Medical Research Methodology*, **12** (1), art. 111.
- Grant, M.J. and A. Booth (2009), 'A typology of reviews: an analysis of 14 review types and associated methodologies', *Health Information & Libraries Journal*, **26** (2), 91–108, doi:10.1111/j.1471-1842.2009.00848.x.
- Greenhalgh, T. (2019), *How to Read a Paper: The Basics of Evidence-Based Medicine and Healthcare*, 6th edn, Chichester: John Wiley & Sons.
- Hawker, S., S. Payne, C. Kerr, M. Hardey and J. Powell (2002), 'Appraising the evidence: reviewing disparate data systematically', *Qualitative Health Research*, **12** (9), 1284–99.
- Higgins, J.P., D.G. Altman and J.A.C. Sterne (eds) (2011a), 'Chapter 8: Assessing risk of bias in included studies, in J.P.T. Higgins and S. Green (eds), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*, The Cochrane Collaboration, Chichester: Wiley Blackwell, pp. 205–28. (The current handbook is available from www.cochrane-handbook.org, accessed 3 March 2020.)
- Higgins, J.P., D.G. Altman, P.C. Gøtzsche, P. Jüni, D. Moher, A.D. Oxman, et al. (2011b), 'The Cochrane Collaboration's tool for assessing risk of bias in randomised trials', *BMJ*, **343** (October), d5928.
- Jeanfreau, S.G. and L. Jack (2010), 'Appraising qualitative research in health education: guidelines for public health educators', *Health Promotion Practice*, **11** (5), 612–17, doi:10.1177/1524839910363537.
- Katrak, P., A.E. Bialocerkowski, N. Massy-Westropp, V.S. Kumar and K.A. Grimmer (2004), 'A systematic review of the content of critical appraisal tools', *BMC Medical Research Methodology*, **4** (September), art. 22, doi:10.1186/1471-2288-4-22.
- Keen, C. and M.E. Otter (2014), 'Critical appraisal', in D. Walker (ed.), *An Introduction to Health Services Research*, London: Sage, pp. 33–50.
- Keller, M.B., N.D. Ryan, M. Strober, R.G. Klein, S.P. Kutcher, B. Birmaher, et al. (2001), 'Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial', *Journal of the American Academy of Child & Adolescent Psychiatry*, **40** (7), 762–72.
- Kmet, L.M., R.C. Lee and L.S. Cook (2004), 'HTA Initiative # 13: standard quality assessment criteria for evaluating primary research papers from a variety of fields', Alberta Heritage Foundation for Medical Research, Edmonton, AB.
- Kuper, A., L. Lingard and W. Levinson (2008), 'Critically appraising qualitative research', *BMJ*, **337** (August), a1035.
- Le Noury, J., J.M. Nardo, D. Healy, J. Jureidini, M. Raven, C. Tufanaru et al. (2015), 'Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence', *BMJ*, **351** (September), h4320.
- Letts, L., S. Wilkins, M. Law, D. Stewart, J. Bosch and M. Westmorland (2007), *Critical Review Form – Qualitative Studies (Version 2.0)*, Hamilton, ON: McMaster University.
- Long, A.F., M. Godfrey, T. Randall, A. Brettell and M.J. Grant (2002a), 'HCPREDU evaluation tool for mixed methods studies', University of Leeds, Nuffield Institute for Health, Leeds, accessed 3 March 2020 at http://usir.salford.ac.uk/13070/1/Evaluative_Tool_for_Mixed_Method_Studies.pdf.
- Long, A.F., M. Godfrey, T. Randall, A. Brettell and M.J. Grant, (2002b), 'HCPREDU evaluation tool for quantitative studies', University of Leeds, Nuffield Institute for Health, Leeds, accessed 3 March 2020 at http://usir.salford.ac.uk/12969/1/Evaluation_Tool_for_Quantitative_Research_Studies.pdf.
- Moon, J. (2005), 'We seek it here . . . a new perspective on the elusive activity of critical thinking: a theoretical and practical approach', Higher Education Academy, University of Bournemouth.
- National Institute for Health and Care Excellence (NICE) (2012), 'Methodology checklist for qualitative stud-

- ies', September, accessed 3 March 2020 at <https://www.nice.org.uk/process/pmg4/chapter/appendix-h-quality-appraisal-checklist-qualitative-studies>.
- O'Brien, B.C., I.B. Harris, T.J. Beckman, D.A. Reed and D.A. Cook (2014), 'Standards for reporting qualitative research: a synthesis of recommendations', *Academic Medicine*, **89** (9), 1245–51, accessed 3 March 2020 at http://www.mmcri.org/deptPages/core/downloads/QRIG/Standards_for_Reporting_Qualitative_Research__A_990451.pdf.
- Pluye, P., M.P. Gagnon, F. Griffiths and J. Johnson-Lafleur (2009), 'A scoring system for appraising mixed method research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Methods Reviews', *International Journal of Nursing Studies*, **46** (4), 529–46, accessed 3 March 2020 at <http://www.ncbi.nlm.nih.gov/pubmed/19233357>.
- Pluye, P., E. Robert, M. Cargo, G. Bartlett, A. O'Cathain, F. Griffiths, et al. (2011), 'Proposal: a mixed methods appraisal tool for systematic mixed studies reviews', accessed 3 March 2020 at <http://mixedmethodsappraisaltoolpublic.pbworks.com>. (Archived by WebCite® at <http://www.webcitation.org/5tTRTc9yJ>, accessed 3 March 2020.)
- Shea, B.J., B.C. Reeves, G. Wells, M. Thuku, C. Hamel, J. Moran et al. (2017), 'AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both', *BMJ*, **358** (September), j4008.
- Specialist Unit for Review Evidence (SURE) (2013), 'Questions to assist with the critical appraisal of qualitative studies', accessed 3 March 2020 at https://www.cardiff.ac.uk/__data/assets/pdf_file/0006/212775/SURE_Qualitative_checklist_2015.pdf.
- Spencer, L., J. Ritchie, J. Lewis and L. Dillon (2003), *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*, London: Cabinet Office.
- Thomas, B.H., D. Ciliska, M. Dobbins and S.A. Micucci (2004), 'A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions', *Worldviews on Evidence Based Nursing*, **1** (3), 176–84.
- Tong, A., K. Flemming, E. McInnes, S. Oliver and J. Craig (2012), 'Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ', *BMC Medical Research Methodology*, **12** (1), 181, accessed 3 March 2020 at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3552766/> and <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-181>.
- Tong, A., P. Sainsbury and J. Craig (2007), 'Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups', *International Journal of Qualitative Health Care*, **19** (6), 349–57.

OTHER USEFUL RESOURCES

- MRC Complex Intervention Guidance: <https://mrc.ukri.org/documents/pdf/developing-and-evaluating-complex-interventions/> (accessed 3 March 2020).
- Vitae Researcher Development Framework: <https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework-rdf-vitae.pdf/view> (accessed 3 March 2020).
- Dr Ben Goldacre's blog: www.badsicence.net/about-dr-ben-goldacre/ (accessed 3 March 2020).
- Sense about Science is an organization dedicated to public engagement with science and disseminating the accurate meanings of published evidence: <http://www.senseaboutscience.org/> (accessed 3 March 2020).
- Study 329:
- Keller, M.B., N.D. Ryan, M. Strober, et al. (2001), 'Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial', *Journal of the American Academy of Child & Adolescent Psychiatry*, **40** (7), 762–72.
- Le Noury, J., J.M. Nardo, D. Healy, J. Jureidini, M. Raven, C. Tufanaru et al. (2015), 'Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence', *BMJ*, **351** (September), h4320.