Notes on 'Appraising the Quality of Research and Evaluation' ROUNDTABLE

Susan Dawe and John Stanwick

Susan Dawe
susan.dawe@ncver.edu.au
John Stanwick
john.stanwick@ncver.edu.au
National Centre for Vocational Ecuation Research (NCVER)

Paper presented at the Australasian Evaluation Society 2004 International Conference 13-15 October-Adelaide, South Australia www.aes.asn.au

Abstract
Participants will discuss one of the key steps in the conduct of a systematic review of research; appraising the quality of research and evaluation. The presenters will introduce the topic from their experience with recently undertaken systematic reviews of vocational education and training (VET) research. Participants will be invited to contribute from their own experience.

Systematic reviews of research are secondary research activities that are aimed at locating all relevant material, both published and unpublished, on a focused policy question. The reviews evaluate content material for information content, approach and robustness, and conclude with a balanced and relevant synthesis of findings. The approach is transparent, making clear the criteria and reasons why studies have or have not been included in the review, and the basis for judgement of quality. These reviews thus provide an empirically based foundation for decision making.

The discussion will assist participants in understanding what is involved in reviewing research in a systematic and objective way. It will also assist those planning and reporting research and evaluation projects.

Introduction

The NCVER has been contracted by the Australian National Training Authority (ANTA) to undertake **a systematic review of research on mature aged workers**. Systematic reviews are a key decision-making tool in many areas of evidence policy and practice internationally. They use explicit and rigorous methods to identify, critically appraise and synthesize relevant research (both published and unpublished) around a specific research question. To the best of our knowledge this is the first systematic review conducted in Australia in the field of vocational education and training. As part of their contract NCVER will be developing a replicable framework and infrastructure within which further systematic reviews of research can be conducted. A steering committee with representation from States and Territories, the Australian Government, ANTA and the research community has been established to oversee this process. The final review question was:

What evidence is there that skill development activities for the mature aged (over 45 years) lead to

(i) improved attachment to the labour market
(ii) improved productivity?

In addition, NCVER developed the following question after discussions with the Australian Indigenous Training Advisory Council (AITAC) to be addressed by a systematic review:

For Indigenous people, what are the key features required in the planning, design and delivery of VET and ACE learning programs to ensure positive educational, employment and social outcomes?

The final reports of these reviews will be available soon from the NCVER website: www.ncver.edu.au

To receive an email when these publications are available, you can subscribe online to NCVER News

What is a systematic review?

The Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre notes that:

> A systematic review is a piece of research. Like any piece of research, it uses research methods that aim to make it produce valid and reliable results. For example, systematic reviews include efforts to find as much as possible of the research which addresses the review's research question. This is important if the review's conclusions are not to be over-influenced by studies which are simply the easiest to find (usually published research). Another example of the methodological approach of a systematic review is the use of a set of explicit statements, called inclusion criteria, to assess each study found to see if it actually does address a review's research question.
>
> As is the case for any good research, the methods for a systematic review are made explicit in a 'protocol' before it starts. This helps to reduce bias in the review process, for example, by ensuring that reviewers' procedures are not overly influenced by the results of studies they find. If changes are needed to the protocol as the review progresses these needed to be noted in the review's final report and the rationale for making changes made clear.
>
> A systematic review is also explicit in reporting its methods so that these can be appraised. For example, the methods used to find studies (database searches, searches of specialist bibliographies, hand-searching of likely journals, attempts to track down unpublished research) will be reported in some detail. This allows readers to decide for themselves whether the reviewers have looked carefully enough to be able to say they have identified as many as possible of the studies that could help answer the review's research question. It is now standard practice for reports of systematic reviews to have clearly defined methods and results sections.
>
> An important characteristic of a systematic review is that it includes a synthesis of its results, which in this case are results from previous research. As a very important part of the synthesis process, systematic reviewers assess the quality of the studies they have found. They can then use this assessment to assign different weights to study findings. Poor quality studies are sometimes downgraded in importance or excluded from the review. The ultimate effect of this is that research can influence a review's conclusions only when that research is sound.
>
> The synthesis is usually presented in the form of a structured narrative, summary tables or a statistical combination (meta-analysis). This synthesis is then used to formulate conclusions and recommendations. The aim is to make the links between the detail of the studies found and the reviewers' conclusions clear.
>
> (The EPPI-Centre is part of the Social Science research Unit, Institute of Education, University of London)

How do systematic reviews differ from literature reviews

1. They are pieces of research in their own right using explicit and transparent methods and follow a standard set of stages. This enables them to be replicated.
2. The process allows for different studies to be weighted for quality and relevance for a given review question.
3. The process produces a map of evidence which helps classify the research.
4. The process allows for reviews to be updated – even by different authors – and so provide flexibility and value for money in the longer term.

5. The review process is designed to support user engagement e.g. practitioners taking part in undertaking reviews.
6. Some systematic review methods enable qualitative and quantitative studies to be analysed and compared in the same review
7. Participating in a systematic review helps improve research skills and can help researchers address how they report on primary research
8. The systematic review process is criterion-based, transparent and public
9. Systematic reviewing enables international collaboration and supports inclusion of international evidence in a review.

Source: 'Systematic Literature Reviews in Education: Advice and Information for Funders', available at http://www.nerf-uk.org/funders/systematic/

Background to Relevance and Quality Criteria

Once the relevant studies have been selected, they will be evaluated by trained reviewers applying explicit criteria in an unbiased way. To ensure the process is transparent, repeatable and meets quality assurance requirements, generic questions are used to extract information in a standardised format (for example, using a standard form, and with coded responses).

Possible criteria

1. The EPPI-Centre guidelines, which set out generic questions relating to the quality of the research, are not available on their website at present as they are being revised. In the interim, NCVER has taken a set of questions from a final systematic review report (as shown in attachment A). The questions cover eight criteria for assessing the quality of any type of educational research. These criteria are:

- adequacy of the description of the context of the study
- sufficiency of the justification for the way the study was conducted
- clarity of the reporting of the aims of the study
- adequacy of the description of the sample used in the study and how it was recruited
- adequacy of the description of the methods used for data collection and analysis
- sufficiency of attempts made to establish the reliability and validity of data collection tools
- sufficiency of attempts made to establish the reliability and validity of data analysis tools
- sufficiency of original data included on terms of enabling mediation between data and interpretation.

The weight of evidence is also determined for three dimensions:

- soundness of method,
- appropriateness of study design and analysis, and
- relevance of the topic focus of the study to the review question

2. An alternative evaluation framework, sourced from Government Chief Social Researcher's office in London, consists of 18 appraisal questions as shown in attachment B.

3. An example from the NCVER's first template for reviewers and the accompanying guidelines for reviewers are shown in Attachment C .

Attachment A

Examples of EPPI-Centre Questions for Quality Criteria and Weight of Evidence

Quality of the study - Reporting
Is the context of the study adequately described?
Are the aims of the study clearly reported?
Is there an adequate description of the sample used in the study and how the sample was identified and recruited?
Is there an adequate description of the methods used in the study to collect data?
Is there an adequate description of the methods of data analysis?
Is the study replicable from this report?
Do the authors avoid selective reporting bias? (e.g. do they report on all variables they aimed to study as specified in their research question or aims?)
Weight of evidence -How representative was the achieved sample (as recruited at the start of the study) in relation to the aims of the sampling frame? Response scale: High/Medium/Low

Quality of the study - Methods and data
Are there ethical issues about the way the study was done?
Were students and/or parents appropriately involved in the design or conduct of the study?
Is there sufficient justification for why the study was done the way it was?
Was the choice of research design appropriate for addressing the research question posed?
Have sufficient attempts been made to establish the reliability of data collection methods and tools?
Have sufficient attempts been made to establish the validity of data collection methods and tools?
Have sufficient attempts been made to establish the reliability of data analysis?
Have sufficient attempts been made to establish the validity of data analysis?
To what extent are the research design and methods employed able to rule out any other sources of error or bias which would lead to alternative explanations for the findings of the study?
How generalisable are the study results?

Weight of evidence A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question? Response scale: High/Medium/Low
Have sufficient attempts been made to justify the conclusions drawn from the findings so that the conclusions are trustworthy? Response scale: High/Medium/Low
In light of the above, do the reviewers differ from the authors over the findings or conclusions of the study?

Review-specific weight of evidence
Weight of evidence B: Appropriateness of research design and analysis for addressing the question or sub-questions of this specific systematic review? Response scale: High/Medium/Low

Weight of evidence C: Relevance of particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question or sub-questions of this systematic review. High/Medium/Low

Overall weight of evidence D: Taking into account quality of execution, appropriateness of design and relevance of focus, what is the overall weight of evidence this study provides to answer the question of this specific systematic review? Response scale: High/Medium/Low

Source: EPPI-Centre: A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills. Reference details: Cifuentes L.& Yi-Chuan J. (2000) Concept Learning through Image Processing. Available at: http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/&Control=ViewItem&Item_ID=IT11325

Attachment B

This evaluation framework was designed to aid informed judgement of quality, and is based around four guiding principles – that research should be:

- contributory in advancing wider knowledge or understanding
- defensible in design by providing a research strategy which can address the research or evaluation questions addressed
- rigorous in conduct through systematic and transparent collection, analysis and interpretation of qualitative data
- credible in claim through offering well-found and plausible arguments about the significance of data generated.

Using these guiding principles, 18 appraisal questions were identified for the assessment framework. These questions cover all of the key features and processes involved in qualitative inquiry—findings, design, sampling, data collection, analysis, reporting, reflexivity and neutrality, ethics and auditability. The appraisal questions suggested are:

1. How credible are the findings?
2. How has knowledge or understanding been extended by the research?
3. How well does the study address its original aims and purpose?
4. How well is the scope for drawing wider inference explained?
5. How clear is the basis of the appraisal of the findings?
6. How defensible is the research design?
7. How well defended are the sample design/target selection of cases/documents?
8. How well is the eventual sample composition and coverage described?
9. How well was the data collection carried out?
10. How well has the approach to, and formulation of, analysis been conveyed?
11. How well are the contexts of data sources been retained and portrayed?
12. How well has diversity of perspective and content been explored?
13. How well has detail, depth and complexity (i.e. richness) of the data been conveyed?
14. How clear are the links between data, interpretation and conclusions i.e. how well can the route to any conclusions be seen?
15. How clear and coherent is the reporting?
16. How clear are the assumptions/theoretical perspectives/values that have shaped the form and output of the research or evaluation?
17. What evidence is there of attention to ethical issues?
18. How adequately has the research process been documented?

For each appraisal question, a series of possible features for consideration in the assessment of quality (quality indicators) are proposed.

Source: Spencer, L. Richie, J, Lewis, J, and Dillon, L 2003, **Quality in Qualitative Evaluation: A framework for assessing research evidenc**e, Government Chief Social Researcher's Office, London. Available at: http://www.number-10.gov.uk/su/qual/index.htm [Contents page]

http://www.number-10.gov.uk/su/qual/ex_summary.htm [Executive summary]

http://www.number-10.gov.uk/su/qual/pdf.htm [Complete report]

Attachment C
Example from NCVER's first Reviewer's template

9. EVALUATION SECTION

In this section we are interested in your evaluation of this study. Please refer to the Reviewers' Guidelines for Quality Appraisal for explanatory detail.

9.1 WEIGHT OF EVIDENCE A (relevance)

Q9.1 How would you rate the **relevance of the particular focus of the study** for addressing the question of this review? Consider the population, intervention and outcomes as described in the review framework compared with those covered in this study. Rate each component as well as giving an overall rating.

In this research study:	Rating				
Population i.e. mature aged	High	Medium+	Medium	Medium-	Low
Intervention i.e. skill development activities	High	Medium+	Medium	Medium-	Low
Outcomes i.e. improved attachment to the labour market or productivity	High	Medium+	Medium	Medium-	Low
Overall Weight of Evidence A (relevance)	High	Medium+	Medium	Medium-	Low

Please add any comment below:

9.2 WEIGHT OF EVIDENCE B (quality)

Q9.2 How would you rate the **quality of this study** in terms of the trust that can be put into its findings against the questions posed? Please refer to the Reviewers' Guidelines for Quality Appraisal.

In this research study:	Rating				
Is the evidence valid?	High	Medium+	Medium	Medium-	Low
Is the evidence reliable?	High	Medium+	Medium	Medium-	Low
Is the evidence authentic?	High	Medium+	Medium	Medium-	Low
Is the evidence sufficient?	High	Medium+	Medium	Medium-	Low
Is the evidence current today?	High	Medium+	Medium	Medium-	Low
Overall Weight of Evidence B (quality)	High	Medium+	Medium	Medium-	Low

Please add any comment below:

9.3 WEIGHT OF EVIDENCE SUMMARY

Q9.3 Please provide final weights of evidence A and B below:

Weight of evidence A | **Weight of evidence B**

Item:	**Weight of evidence A**
Criterion A:	**How would you rate the relevance of the substance of this study to the review question?**
Guidelines:	Consider the following in answering the question above: 1. Actual population sample covered compared to that specified in the review framework 2. Actual intervention(s) covered compared to that specified in the review framework 3. Actual outcomes focussed on compared to those specified in the review framework
Item:	**Weight of evidence B**
Criterion B:	**How would you rate the quality of the study in terms of trust that can be put into its findings for the research question posed?**
Guidelines:	Consider all quality criteria as outlined below ie validity, reliability, authenticity, sufficiency and currency
Criterion 1:	**Is the evidence in this study valid?**
Guidelines:	Consider the following in answering the question above: 1. research aims and variables or concepts measured 2. design and whether methods measure what was intended to be measured (i.e. validity) 3. efforts made to address the validity of data collection tools/methods eg pilot testing tools 4. efforts made at data analysis stage to address validity e.g. limiting analyses where numbers are insufficient 5. efforts made in the analysis to control for bias from confounding variables 6. whether links between data, interpretation and conclusions are valid 7. justification of conclusion drawn
Criterion 2:	**Is the evidence in this study reliable?**
Guidelines:	Consider the following in answering the question above: 1. efforts to determine that data collection methods and tools will yield same result each time (i.e. are reliable) 2. efforts to ensure data analysis can be repeated and yields same result each time 3. any assumptions/theoretical perspectives that shape the form or the output of the research 4. any alternative explanations for stated findings
Criterion 3:	**Is the evidence in this study authentic?**
Guidelines:	Consider the following in answering the question above: 1. whose voice it is in the report 2. for whom and for what purpose the knowledge was being sought 3. sources of evidence: direct or indirect

Criterion 4:	**Is the evidence in this study sufficient?**
Guidelines:	Consider the following in answering the question above: 1. sample sizes etc 2. the authors' conclusions 3. whether there are any other possible explanations for the findings 4. that the evidence presented is enough to support the findings and conclusions
Criterion 5:	**Is the evidence in this study current today?**
Guidelines:	Consider the following in answering the question above: 1. when the study was done and any contextual issues that are no longer relevant or that now are relevant that were not at the time of the study

The weight of evidence A & B will then be scored in a 5 x 5 matrix:

	Weight of evidence B (Quality)				
Weight of evidence A (Relevance)	**H**	**M+**	**M**	**M-**	**L**
H	HH	HM+	HM	HM-	HL
M+	M+H	M+M+	M+M	M+M-	M+L
M	MH	MM+	MM	MM-	ML
M-	M-H	M-M+	M-M	M-M-	M-L
L	LH	LM+	LM	LM-	LL