AAEE 2012 CONFERENCE
Melbourne, Australia
www.aaee.com.au/conferences/2012/

Quality in Engineering Education Research: arriving at consensus

LesleyJolly[a], Keith Willey[b] and Anne Gardner[b]
Strategic Partnerships[a] University of Technology Sydney[b] ,
Corresponding Author Email: ljolly@bigpond.net.au

BACKGROUND

Arguably the most important opportunity to acquire the standards and norms of any discipline and develop researchers' judgement is the peer review process – and this is probably particularly true in an emerging discipline such as engineering education. Ironically, research in many disciplines has established that the review process is deeply flawed in conception as well as (often) in operation, with the American Medical Association asserting that if peer review were a drug it would never be allowed on to the market. And yet university ranking systems for published research, on which all of our careers depend, rely on this flawed instrument. With this in mind we have been examining how members of our community (AAEE) give and respond to reviews with a view to making the process more useful.

PURPOSE

Reviewing is an inexact and subjective process so it would be misguided to think that somehow inter-rater reliability or some notion of objective 'truth' may be attained. Instead, we ask what reviewers need to do to provide helpful advice that can help shape norms and standards in the field.

DESIGN/METHOD

In previous work (Willey et. al. 2011; Jolly et.al. 2011) there appeared to be a need for well-expressed criteria that would guide authors on what a publication should contain and guide reviewers in how judgements should be made. With the help of a Delphi panel made up of 12 international researchers in the field a set of criteria were developed. Volunteers were then sought to apply the criteria to sample texts in an online tool (SPARK[plus]). Individual interviews with some respondents were then used to clarify participant's understandings and goals.

RESULTS

The criteria developed by the Delphi panel are those being used for this conference. The members of the panel particularly approved the 'comments' accompanying the criteria *per se* which were intended primarily as guidance to authors about acceptable practice. Anecdotal evidence to date suggests that authors should find the criteria and comments clarify expectations but the matter of standards will remain. The use of the criteria in the second stage and analysis of the discussions in particular will produce information both about present expectations and practices and visions of future growth and improvement.

CONCLUSIONS

Our analysis of the stage 2 data will aim to describe consensus on research quality and how to use the peer review process to help attain it, in the form of recommendations for future application of the criteria, in journals as well as at conferences. Our international experts from the Delphi panel have expressed an interest in being involved in stage 2 and informed about the outcomes so the potential also exists for this community to develop best practice peer review in engineering education thorough the sharing of their expertise in this way.

KEYWORDS

peer review, research quality, career development

Introduction: The Problem with Reviews

Engineering Education research is still a developing field of intellectual endeavour, beset by difficulties of competing epistemologies and methodological assumptions (Case and Light 2011) and as a result norms are not well established and the quality, focus and even purpose of work in the field remains variable. While it would be unrealistic and counter-productive to look for complete agreement between reviewers (Fitzpatrick 2010, Lipworth and Kerridge 2011), we have elsewhere argued (Jolly et. al. 2011 and Willey at al 2011) that the review process ought to be a place where expectations could be clarified and authors helped through adjustment to a different kind of research than the one they were trained for. However, that assumes that reviewers are able and willing to articulate what good engineering education research would look like, rather than rely only on their own unexamined assumptions built on experience in quite different kinds of work. In engineering education we see this where reviewers reject a piece of work written in an unfamiliar format or using unfamiliar methods with the words “as an engineer…”. Education is a different kind of undertaking from engineering and engineering education research needs to establish an epistemology and focus that generates new insights. Considering how reviewing shapes disciplines, Origgi (2010) calls for “epistemic vigilance”; a reflexive and critical stance on the reasons, biases and pressures that make some topics [and methodologies] emerge and thrive and not others. That is to say, a review that can help an author and help develop professional norms in the field needs to be explicit about its assumptions, knowledgeable about the field and open to argument. Only thus can we avoid painting ourselves into a corner where we go round and round the same topics using the same methodologies and not producing interesting and insightful arguments. Although the standard blind review process does not foster academic argument and the progression of ideas (Fitzpatrick 2010), even within this process, it ought to be possible for reviews to have some positive impact on the development of the field. Sadly, our analysis (Willey et. al. 2011) of recent AAEE conference review processes suggests that we are not moving towards this goal.

Results of prior work

In previous years (Jolly et. al. 2011, Willey et. al. 2011) we have analysed reviews and authors’ responses to reviews to identify what reviewers are focussing on as important in a paper, and thus making central to the field, and what use authors are able to make of the feedback they get. Tables 1 and 2 refer to our analyses of 66 reviews and 23 author responses from the AAEE2010 conference.

Table 1: Types of response given by reviewers (from Jolly et. al. 2011)

Type of response	No. of reviews
Gaps in logic	39 mentions (59% of reviews)
Typographical/grammar errors	34 mentions (51% of reviews)
Inadequate data handling	27 mentions (41% of reviews)
Identifying relevant literature	17 mentions (26% of reviews)

Table 2: Categories of author assessment of their reviews (from Jolly et. al. 2011)

Category	No. of author responses
Counterproductive	3
Ineffectual	14
Cruel to be kind	1
Positive Benefit	10

We have argued elsewhere (Jolly et. al. 2011 and Willey at al 2011) that, as is widely reported for other disciplines, the review process in engineering education is failing in the goal of affecting author behaviour since only 10 of the responding authors found their reviews helpful in any way. We suggested that this may have been because reviewers were concentrating on aspects of the work that weren't crucial or that they weren't giving sufficiently specific advice for authors to make the necessary improvements. One thing that we suggested may be of value in making reviews more helpful was to clarify expectations through the use of explicit and detailed criteria.

As a result of this work we put a proposal to the executive committee of AAEE that we do further work to develop internationally recognised criteria for papers in engineering education and begin activities to disseminate these criteria through the community and begin to develop appropriate standards. The executive and the 2012 conference organisers have been helpful in supporting this work and this paper reports on it.

Methodology

It was proposed to use a Delphi Panel to arrive at a set of criteria which would then ask members of the community to volunteer to try out on some short mock papers. Online discussion amongst this cohort of reviewers was then expected to help cement standards and indicate possible direction for future work. In the event, this methodology was not able to be implemented exactly as planned.

Delphi Panel

A Delphi panel is a standard research technique where consensus is required on recommendations or findings (DuBois and Deuker 2009). The "experts" may be anyone with an informed interest in the topic at hand and ideally should come from a range of contexts. Our 12- person panel included 3 academics from the UK and USA who are widely published in engineering education as well as Australian participants with varying degrees of expertise and reputation in the field. The less experienced members were included to ensure that we were not just repeating the views of an established 'clique', in the spirit of "epistemic vigilance". In the event there was broad general agreement amongst the panel members and we needed only two iterations of the proposal to arrive at consensus.

The first round proposal drew on the criteria used by the Journal of Engineering Education as it is the highest rated journal in the field and thus provides a kind of 'gold standard'. Nevertheless, some modifications were felt necessary to their criteria, and so panel members were asked to comment on these changes. Figure 1 illustrates the process for the first criteria.

CRITERION 1

The first JEE criterion reads:

state clearly the questions or propositions addressed and the significance of the research to engineering education research or practice (*focus and relevance*);

Recently it has been suggested that another way of articulating this point is:

Briefly describe the research question/hypothesis/focus of the change in your practice, and the rationale behind this.

Which form of words do you think will be most effective in helping authors to articulate what their research/paper is about and why it matters? Why?
Type here

We have noted that many authors seem preoccupied with the local significance of their work only. That is, how it helps them to solve a problem in a particular course or institution. Is there a need to explicitly prompt them to refer to the wider significance, and if so, how is it best done?
Type here

Figure 1 Round 1 Delphi process

Once their responses from all of the panel members had been collated and the criteria rewritten in light of their comments, a second version was circulated for further comment. It was this version that was agreed upon and adopted for the 2012 conference (Figure 2)

OVERVIEW OF THE CRITERIA
The following list of criteria was generated by a Delphi Panel of national and international experts specifically for use with
the AAEE annual conference as well as potentially in AAEE's *Australasian Journal of Engineering Education*. The comments in italics are intended as clarification for authors and reviewers and form an important part of the criteria.
1. FOCUS OF THE PAPER
The text clearly:
• describes the research question / hypothesis / focus of the study; and
• explains the implications of the project for engineering education research or practice.
"Implications" may include consideration of whether the findings could be applied elsewhere, how the findings might be
confirmed or refuted in other contexts, or how the work reflects on existing literature in the field.
2. RELEVANCE
The text clearly:
• relates the work undertaken to relevant discussions in the engineering education literature and other disciplinary literature as required; and
• describes its contribution to these discussions.
The contribution may be confirming, building on or challenging existing work, or contributing to new theory.
3. METHODOLOGY
The text clearly:
• describes and justifies the appropriateness of the overall design, methods, theories and analytic processes; and
• discusses the limitations of the study.
"Appropriateness" means appropriate to the goals and questions of the study and the argument being pursued.
"Limitations" refers to the limitations of the study and not of the methodology as whole.
4. ARGUMENT
The text clearly presents original ideas or results of general significance that are:
• supported by convincing evidence, and
• clearly reasoned, illustrating the connection between claims and evidence.
Evidence is made convincing by demonstrating that the kinds of evidence used are appropriate to the question being asked and the way it supports the argument.
5. WRITING QUALITY
The text uses appropriate English language of a sufficient standard to clearly convey the argument and enable the reader to make sense of the paper.
Authors are encouraged to have a colleague read the paper before submission, especially where English is not their first language.
6. USE OF ILLUSTRATIONS
Uses Tables and Figures only where they clarify the argument, and Tables and Figures are meaningfully explained in the text.

Figure 2: Final criteria

Test reviews and interviews

Invitations were sent to 300 past AAEE authors and this year's list of reviewers, asking them to read their choice of four short papers and rate and comment on them using the criteria. The papers were created by us by severely cropping existing texts (with the authors' permission). The online tool being used allowed us to release all the reviews and ratings anonymously and we had hoped to organise online discussion groups where participants could discuss such issues as the range of responses, how well the criteria meted the authors' needs and how the presence of the criteria affected their judgement.

In the event, the response rate was disappointingly low with no paper gaining more than six reviews and instead of trying to have an online focus group we asked for volunteers for individual interviews. Four interviews were held. The focal topics were:

- the reviewer's previous experience of reviewing and how they went about the task generally,
- how they found reviewing to these criteria and
- what they thought of other reviewers' responses.

These numbers are clearly too small to draw wide conclusions from but they do raise interesting questions for further investigation and particularly when compared with our previous work on this topic.

Findings & Discussion

Table 3 shows the breakdown of the types of response provided by reviewers in this sample compared to the previously analysed years.

Table 3: Types of response in test reviews.

Type of response	No. of reviews
Gaps in logic	33% of reviews (59% previously)
Typographical/grammar errors	39% of reviews (51% previously)
Inadequate data handling	66% of reviews (41% of reviews)
Identifying relevant literature	0% of reviews (26% of reviews)

The high levels of attention to inadequate data handling may be a result of the way we edited the papers to make them short enough for inclusion in the exercise. There was still some concern with typographical errors even though the criteria seek to de-emphasise this aspect of reviewing. There were no instances in our sample of the kind of advice that identifies a lack in the paper AND tells the authors where they may go in the literature for more information or a model for analysis or any of the other directly helpful instructions that authors have previously told us they were looking for.

Another striking aspect of these comments was the way in which they were made explicitly in light of the reviewer's own personal preference or engineering training, rather than the needs and expectations of educational research. For instance the reviewer who told authors "I prefer tables" may well have good reasons for doing so but it does not help an author to know that unless they are also told what this preference is based on. Similarly, the reviewers who refer to qualitative data as "*someone's 'story'*" or "*anecdotal evidence*" are applying the standards of engineering rather than contributing to the establishment of norms and standards in the emergent field of engineering education. While most reviewers and interviewees in our study agreed that papers need "*data and analysis to support an argument*" there was still some discussion needed in the community around what we accept as data. Origgi (2010) warned against this behaviour when she called for epistemic vigilance, as it impairs the development of new ideas and approaches, but may just reflect the difficulties engineers have with a new research paradigm (Borrego, 2007).

Engineering academics' preparedness to take on different methodologies may also be related to how well they can generalise about their own research and practice processes. For example engineers are supposed to be good problem solvers. Good practice in problem solving is to use a solution method that best suits the problem you're trying to solve – this involves appraising the problem from a range of perspectives and being open to alternative solutions that you haven't used before rather than using one of a small range of methods that you are familiar with.

Many journals and conferences use a rating scale from -3 to +3 (representing Strong Reject, Reject, Weak Reject, Borderline Paper, Weak Accept, Accept, Strong Accept) either for papers as a whole or for separate criteria. In this exercise reviewers were asked to rate each paper on such a scale for each criterion and then add at least five words justifying their rating. Our program analysed the ratings as ±within each discreet rating label so as to account for its use of the continuous scale and hence the difference between for instance a low Accept and a high Accept,. That is Accept -, Accept and Accept+ represents the lower, middle and upper 33.3% of the Accept rating division respectively. Some reviewers used only the rating scale and did not provide comment. However, a number of inconsistencies in how the scale was used raises questions about the usefulness of such a measure. It is no surprise that there is disagreement between reviewers but the nature of the disagreement was instructive.

For instance, one paper received two reviews, both of which agreed that the paper was lacking in methodology and argument. But one reviewer rated both these criteria as Strong Reject, while the other rated it respectively as Weak Accept and Borderline . How is a busy editor or conference organiser to respond to such divergence. Then there is the review that judges the methodology to be inadequate but still rates the paper against this criteria as Weak Accept or the paper that had no figures or illustrations but still got marks on the final criteria from all but one reviewer. It is hard to see what function ratings such as these can perform in the review process if taken at face value.

The interviews threw some light on the preconceptions and assumptions that reviewers bring to their task when it comes to engineering education and result in the patterns described above.

Criteria, implicit and explicit

All of our interviewees told us that they undertook the onerous task of reviewing (and contributed to our project) in order to contribute to the discipline and community. This indicates to us that we are not the only ones who expect reviews to be influential. All of them had previous experience in reviewing in their specialised disciplinary areas and it appears that none of those disciplines provide criteria for authors' or reviewers' use but rely on implicit professional judgement. There may be justification for this in well-established disciplines but it does not encourage reflection on epistemological and methodological issues. Our four interviewees had various levels of experience in engineering education research but all of them confirmed that they built their judgements in this field on previous experience and personal factors. As one said "reviewing is a personal experience…would I recommend it [the paper] to a colleague?" Others admitted to not having extensive experience in educational research but were still prepared to make judgements such as two case studies not providing sufficient evidence to support an argument. We suggest this epistemological divide is, at least in some part, due to the high level of consensus in many specialist technical areas compared to the educational domain eg many of us work within the constraints of physical laws. This high level of consensus and repeatability of both our methods and results means that we don't have to explain our perspective and argue the logic of our method as much as in the social sciences. We will describe our method but are generally not required to expend much effort articulating our reason for choosing it or for explaining why it is appropriate. For example, if one was testing structural elements they would indicate where they located the strain gauges on the test specimen, but wouldn't have to explain why using strain gauges were the best way to measure the phenomenon that they were investigating. In the educational domain there are many theories (not laws) and a variety of perspectives to view any learning context, researchers described using a particular theory as a lens through which to analyse the data ie the theory or perspective becomes the filter through which the context is viewed.

As mentioned earlier the concept of repeatability of the method, investigation variables and/or process is a major difference between stereotypical engineering research and engineering education research. Educational practices are usually much more sensitive to context than engineering research practices. One of the objectives of the scientific method is that the 'experiment' can be repeated anywhere and the results obtained will be the same, within acceptable tolerances. There is no such objective in educational research. Studies are usually about understanding a particular context. That context needs to be comprehensively described so that readers can begin to understand the associated issues, and to discern if the context shares attributes with their own situation and assess how transferable it is likely to be.

Consensus on "what counts"

We observed during the Delphi panel stage that there was broad agreement on what criteria should address and how they should rate performance. This consensus was obvious in most of the reviews also, numerical ratings aside. One interviewee told us that when he looked at other people's reviews he was pleased to see that they "agreed on the obvious". However, such consensus may be a problem in this field if it is based on the norms of regular engineering rather than those of fields more used to dealing with educational topics (Borrego 2007). One could argue that it is less important whether the reviewers agreed or not compared to whether they could articulate why they rated the paper as they did. This articulation is what provides the constructive feedback to the author about how to improve their paper. Some of our interviewees were much less confident on what to insist on in relation to methodology in educational related papers compared to papers in their specific discipline area. This didn't stop anyone making a judgement though!

Issues arising from the application of criteria

One of the reasons for devising the criteria is from a developmental perspective in recognition that most engineering academics have not been trained in educational research. As a result, many new to the field tend to write the so-called 'show and tell' or 'Marco Polo' paper when they first submit to the conference. The criteria provide a framework for writing a more 'scholarly' paper by asking authors to, for example, relate their work to existing literature. More experienced authors would know they need to situate their work in relation to prior research so are less reliant on the criteria.

However, it is not just the level of experience of the author that affects the use of the criteria, the type of experience of and how the author sees themselves (for example novice developing or expert researcher) is also relevant. A further limitation occurs when the criteria are combined with a prescribed paper format that often seems to limit the type of acceptable papers. Prescribed conference formats within the engineering education community are often reminiscent of the standard laboratory report that engineers have been writing since their first year of undergraduate studies. This format suits quantitative data but it's more difficult to use with other types of data and ways of creating arguments. For example, the authors when writing this paper found that separating the results and discussion made the argument disjointed and so chose not to do so as prescribed by the structured abstract template. It is difficult for in any prescribed structure to accommodate the range of methods resulting from the interdisciplinary and often more qualitative nature of engineering education research. As one of our interviewees commented the structure "shoe-horned authors into the scientific method." As an exercise we looked at ten educational research journals that were ranked at the A* level when ERA was ranking journals eg Assessment & Evaluation in Higher Education. Of these ten journals eight listed their most cited articles on their website. While the authors applaud the committee's intent, of these eight articles, including papers by such luminaries as John Biggs, David Nicol and Sue Clegg, only two would fit the structure suggested for this year's conference.

These issues with applying guiding criteria and structures to different types of papers reflect the tensions in the AAEE community in regard to the Association's mission and purpose. Are we mainly a practice community existing to exchange teaching tips? Or should we be more focussed on contributing to the foundational knowledge ie research in this field? Does it have to be a binary decision? Or should our objective be to create a 'broad church' that includes people both practice-oriented and research-oriented. Hence, recognising the symbiotic relationship in engineering education between practice and research and acknowledging that practice is frequently the pathway through which many engineering academics initiate research.

Summary

The review criteria developed as part of this project are helpful but need to be applied with care especially when the paper presented is different from the format expected. Qualitative methodologies, theoretical discussions and papers that assume a certain level of expertise all may have to be given some leeway so as not to limit what can be achieved by engineering education research. The current criteria are like the community of AAEE – a 'work in progress'.

References

Borrego M. (2007) Conceptual Difficulties Experienced by Trained Engineers Learning Educational Research Methods. *Journal of Engineering Education*, 96, 2 , 91-102.

Case, J., Light, G. (2011) Emerging methodologies in engineering education research. *Journal of Engineering Education 100*(1): 186 – 210.

Fitzpatrick, J., 2010, "Peer-to-peer Review and the Future of Scholarly Authority" *Social Epistemology 24*(3):161-179.

Jolly, L., Tibbits, G., Willey, K., Gardner, A. (2011) Conferences, reviews and conversations about improving engineering education. Paper presented to REES Symposium, Madrid.

Lipworth, W., Kerridge, I., 2011, "Shifting Power Relations and the Ethics of Journal Peer Review" *Social Epistemology, 25*: 1, 97 — 121

Origgi, G., 2010, "Epistemic Vigilance and Epistemic Responsibility in the Liquid World of Scientific Publications" *Social Epsitemology, 24* (3): 149-159.

Willey, K., Gardner, A., Jolly, L. (2011) Improving research quality through the peer review process. Paper presented to SEFI Conference, Lisbon.

Copyright statement