

Gatekeeping or filtering?: Investigating the connection between peer review and research quality

Keith Willey¹, Lesley Jolly², Gregory Tibbits², Anne Gardner¹

¹University of Technology, Sydney; ²University of Queensland
keith.willey@uts.edu.au; ljolly@bigpond.net.au; gregory.tibbits@uqconnect.edu.au;
Anne.Gardner@uts.edu.au

Abstract: Arguably, the most important opportunity to acquire the standards and norms of the discipline and develop researchers' judgement is the peer review process – but this depends on the quality of the reviews. 'Good' feedback - which we take to mean feedback that has the capacity to improve subsequent practice - has been identified as being timely, specific and relevant. Yet often reviews lack these basic qualities. In this paper we report an investigation of the peer review process at the 2010 Australasian Association of Engineering Education (AAEE) conference. Authors at the conference were given the chance to rate their reviews and we subsequently analysed both the nature of the reviews and authors' responses. Findings suggest that the opportunity to use the peer review process to induct people into the field and improve practice is being missed. As in other disciplines there is also ample evidence that the review process does little or nothing to ensure the standard and relevance of conference presentations. It is therefore legitimate to ask whether there may not be better processes to attain these ends and we conclude with some discussion of how the review process may be made more helpful for everyone involved.

Introduction

If we are to talk of inclusive engineering we must consider our own practices in venues such as this one. What diversity of ideas is able to emerge through the normative practices of setting conference themes, reviewing of papers and the potentially confrontational conference presentation itself? How are neophyte engineering education researchers able to be developed within formal structures that reward the familiar and the well-established? While conference presentation is sometimes regarded as a preliminary step to mature publishing in journals (the gold standard for academic practice), are conferences such as ours really venues that can foster scholars and their ideas? We wish here to raise questions about one aspect of scholarly practice which we usually take for granted, the peer review system, which might actually work to discriminate against innovation and impede the development of the field. While we acknowledge that there are external pressures which are likely to maintain the need for peer review, we would like to open debate on how we might diversify our practices in productive ways.

It has been well demonstrated that there are many ways in which peer review fails (Goodstein 2000) including through restricting the dissemination of new ideas, excessive in-group gatekeeping and inconsistency (Fitzpatrick 2010, Lipworth and Kerridge 2011). Origg (2010) has argued that an important function of the review process is a “conversation in slow motion” amongst scholars in order to improve, disseminate and develop their ideas and we argue that such a conversation is much need in the emerging and as yet immature field of engineering education research. Academics engaging in this field frequently have to argue for the rigour and significance of their work (JEE 2011). While some of the necessity for this argument is a result of institutional resistance to the pursuit of educational research by engineers, some of it arises from a perception (sometimes accurate) that conference presentations and published work in this still emerging field is not of the highest quality. Given that the majority of people who undertake this research were trained in quite different paradigms than those that underpin educational research, bridging the epistemological divide between technical training and the more social science of education continues to be problematic for some (Borrego 2007).

If the quality and rigour of engineering education is to be improved, the existing research community needs to do more to move engineering education researchers from being novice to more expert as they develop their research. This requires not only developing knowledge of the relevant research theories and methods, but for novices to develop their judgement about what is acceptable practice in the field (Case and Light 2011). Arguably, the most important opportunity to acquire the standards and norms of the discipline and develop researchers' judgement is the peer review process (Fitzpatrick 2010) – but this depends on the quality of the reviews. 'Good' feedback has been identified as being timely, specific and relevant (Gibbs and Simpson 2004), yet often reviews lack these basic qualities, perhaps because reviewers see their task as a gatekeeping one rather than quality improvement or development. For example, the authors recently received a review for a research paper investigating peer reviews that stated:

I personally do not think that it is the reviewers's role to be friendly, or provide training to naïve researchers. Most important seems to assist the editor in deciding whether a paper is acceptable or not for publication (i.e. screening).

With the advent of various forms of online publishing the peer review process has come under increasing scrutiny (Fitzpatrick 2010), and here we want to focus on the way that peer review can and ought to be a *conversation* among people with similar intellectual interests for the purposes of improving understanding and expressing (and where necessary altering) community standards and opinions.

For many new scholars, a conference paper can often be a first way to try out an idea on those who can offer informed opinion. Many scholars plan to turn conference papers into journal articles in the light of feedback received on the initial presentation. With increasing numbers of ways to get such feedback (including open review sites on the Web), conferences themselves risk becoming irrelevant to this purpose. With ever decreasing time to present and have discussion at most conferences, the comments of anonymous reviewers of the written paper may be the only developmental help many authors get from conference attendance. It is therefore worth asking just how well conference reviews are performing a developmental role.

The Australasian Association for Engineering Education (AaeE) has been actively pursuing improvement in engineering education research quality, including examination of the quality of the peer review process. At the 2010 conference, authors were offered the opportunity to rate and comment on the reviews their papers had received. Authors were also asked to use an online tool (SPARK^{PLUS} [9]) to rate the quality of their reviews and this de-identified information was discussed with delegates in an open forum at the conference. Since that time we have refined the analysis substantially.

Methodology

In the analysis that follows we consider only the reviews of papers that were accepted into the "Research" category (n= 66 reviews two each for 33 papers), since clearly expressed criteria were provided for these papers and were available for authors and reviewers from the beginning of the review process. These criteria are listed in Figure 1. In this category 23 authors (70%) responded to the invitation to assess and comment on their reviews and we have considered all of these here. The online tool used to collect these assessments comprised seven questions rated on a Likert scale (strongly disagree to strongly agree - see Figure 2) and a set of open-ended questions.

1. Context and research questions	
Excellent	The situation being investigated is clearly and concisely described and generates the research questions in a logical manner. The paper shows evidence of familiarity with the research literature in engineering education and where appropriate more widely. The research questions make clear what the researcher wanted to know about the situation and are questions that can generate valid and reliable answers.
Good	Covers all points above but less clearly and systematically. The description of the situation may lack relevant detail OR the use of existing literature may be sketchy or tokenistic OR the

	questions may lack clear logical connection with the situation and literature and/or may not be susceptible of clear answers.
Poor	More than one of the faults referred to above.
2. Theoretical frameworks	
Excellent	The research is clearly situated within overarching explanatory frameworks appropriate to the research questions and the situation being researched. The concepts of the theoretical frameworks are used to structure the data gathering and/or analysis.
Good	Shaky or partial links between the theory and the research questions and data.
Poor	No or token use of theoretical frameworks or theory that is not well chosen for the research questions. Misunderstood theory.
3. Methodology	
Excellent	The methodology describes the logic of the connection between what the researcher wanted to know and the data gathering process, making clear the strengths and limitations of the methods chosen. Well chosen and imaginative data gathering methods.
Good	Adequate but limited (?pedestrian) choice of methods. Sketchy rationale.
Poor	No rationale for choice of methods. Poorly chosen methods.
4. Findings and Conclusions	
Excellent	The findings are well argued on the basis of the data presented. Alternative explanations are considered and their rejection explained.
Good	Good connection between data and conclusions.
Poor	Findings do not account for all of the data presented or are not well supported by the data.
5. Discussion	
Excellent	This section returns the reader to a consideration of the starting point of the research. It may discuss how the findings clarify the original situation of interest, throw new light on the theoretical stance taken or the methodological adequacy of the research and/or make recommendations for engineering education research more widely. Answers the 'so what?' question.
Good	Points out the relevance of this research for understanding the original situation but makes limited reference to wider application.
Poor	Fails to connect the results of the research with the wider research environment.

Figure 1. Criteria for research papers at AaeE 2010.

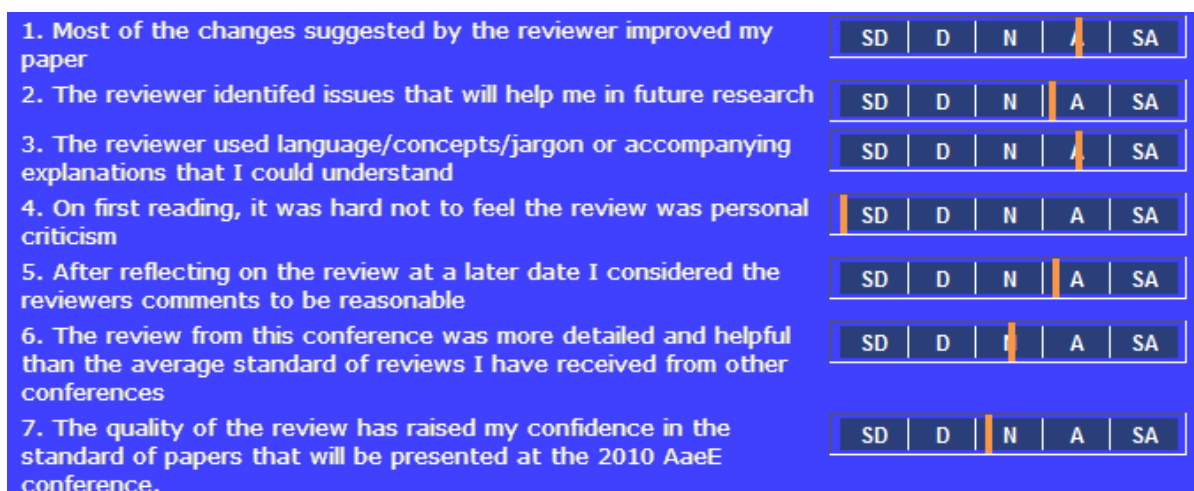


Figure 2. SPARK^{PLUS} screen shot showing the Likert scale (SD strongly disagree to SA strongly agree) questions for authors in regard to the reviews they received for AaeE 2010.

We were surprised by some of the favourable ratings authors awarded their reviews. To benchmark these ratings, the four members of our team read through all 46 of the relevant reviews and rated them ourselves. We also developed a content analysis of the reviews in an attempt to gain a more objective description of them.

Findings

The nature of the reviews

The most striking characteristic of the reviews was the fact that out of all of them only 4 made explicit reference to the criteria and these did not all include a rating against each criterion. A further 3 included some rating either of sections of the paper or the paper as a whole, but these ratings did not necessarily relate to the given criteria. So, for instance, some reviewers used the authors' section headings to structure their review and assigned a rating to each section, whether or not these headings matched the criteria.

Content analysis revealed that most of the advice given in the reviews fell within the five categories summarised in Table 1.

Table 1. Categories of explicit advice given by reviewers AaeE2010

Type of advice	No. of reviews
Gaps in logic	39
Typographical/grammar errors	34
Inadequate data handling	27
Identifying relevant literature	17
Quoted from paper under review	12

Gaps in logic were identified with advice such as *“lack of alignment between the research question, the literature cited, the data and the analysis”* or the fact that *“conclusions are not supported by the data”*. While this is potentially helpful advice, a lack of specificity was often apparent. New scholars may need to be told what exactly alignment looks like for instance.

The attention to typographical errors surprised us but is in line with journal review practice. However, if we are concerned to open debate and improve actual research practice, we argue that a more relevant model for conference reviewers might be the writing circle or collegial responses to drafts, where content is more important than form.

Reviewers comments on data handling tended to be more specific, drawing attention to ways to improve tables or what extra information could be added, for instance. Similarly, where reviewers drew attention to specific sources in the literature we expected that this would help improve not only the paper under discussion but future work as well. There were examples such as *“Make a direct connection to the Bradley Review and the government response, so all readers can understand the importance of this study and how it could relate to their institution”* and others giving actual citations. Quotation from the paper under review was in our opinion another form of specificity in advice but only 18% of the reviews identified the actual words in the paper that were of concern.

Author ratings of reviews

As Figure 3 illustrates, author ratings on the Likert scale show that 63% thought the review had helped improve their paper. Interestingly approximately 61% of responding authors reported that *“On first reading, it was hard not to feel the review was personal criticism”* and even after reflecting on the review at a later date nearly 44% of these authors considered the reviewers' comments to be unreasonable. Given the review characteristics we identified earlier in section 3, it is hard not to think that at least some authors are taking well directed criticism too personally. Figure 4, on the other hand suggests that authors are more unsure about how to apply the review advice to future work. This is an issue if we want reviews such as these to help develop ideas and the field of research.

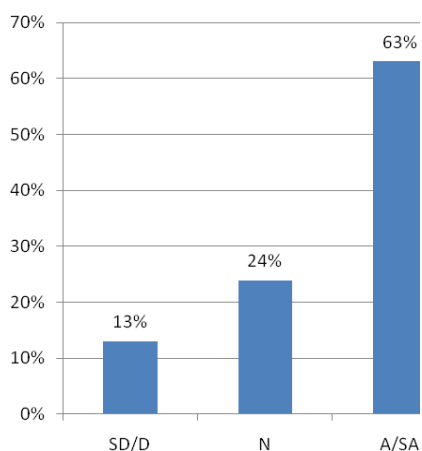


Figure 3. Most of the changes suggested by the reviewer improved my paper .

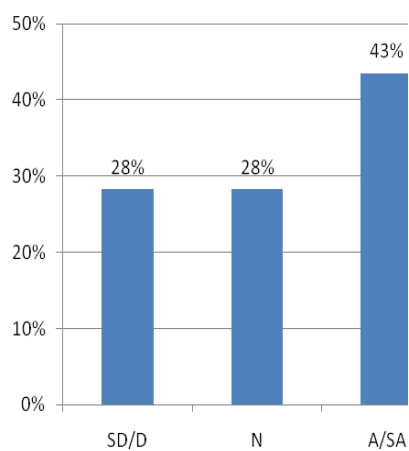


Figure 4. The reviewer identified issues that will help me in future research.

In a further step in our analysis we compare our own ratings of the reviews on these two questions and author ratings and tabulated them against reviewer's self-identified level of confidence.

Table 2: Ratings correlated against reviewer confidence

Research Team Question 1				Authors Question 1			
	Medium Confidence	High Confidence	Expert		Medium Confidence	High Confidence	Expert
D/SD	2.7%	11%	1.6%	D/SD	2.2%	2.2%	2.2%
N	5.5%	18.3%	5.5%	N	2.2%	13.3%	4.4%
A/SA	7.2%	30.5%	17.2%	A/SA	11%	44.5%	17.8%
Research Team Question 2				Authors Question 2			
D/SD	7.8%	34.5%	13.3%	D/SD	2.2%	20%	2.2%
N	5%	10%	4.5%	N	6.6%	15.5%	11%
A/SA	2.7%	15.5%	6.6%	A/SA	6.6%	24.5%	11%

We would like to be able to say with certainty that the level of reviewer confidence correlated with agreement that the review was helpful but the number of cells is too small to confirm this by ANOVA. There does appear to us to be a clear tendency for authors to agree more with reviewers of high rather than expert confidence and we wonder whether those who consider themselves expert are also those whose tone suggests personal criticism.

Open-ended comments on reviews

Not all authors took the opportunity to comment in writing on their reviews, but of those who did were analysed into the categories in Table 3.

Table 3: Categories of author responses to open-ended questions

Category	No. of author responses
----------	-------------------------

Counterproductive	3
Ineffectual	14
Cruel to be kind	1
Positive Benefit	10
Future impact	2
Total	30

Counterproductive reviews were those which authors felt were actual impediments, either because the tone of the review was too carping or the reviewer had misunderstood the work. The high number of ineffectual reviews is particularly worrying in the light of the present discussion calling for reviews to have positive benefit and improve future work. In the eyes of the authors at least, the unhelpful reviews outnumbered the helpful ones.

The way forward

Peer reviewed publication is a process of central importance to scientific inquiry. It is a process that is the standard method of communication allowing researchers to inform others of their work, and a process by which a conversation between researchers occurs (Origgi 2010). Peer review also performs a gatekeeping function, supposedly ensuring quality research is published (Lipworth & Kerridge 2011; Fitzpatrick 2010) and thereby improving the quality of research. The problems we have identified here relate to the adequacy of peer review for these purposes. We especially question the relevance of applying standard norms of peer review to conference presentations which ought, we contend, to foster conversation and debate.

One model that has been suggested elsewhere derives from social media (Fitzpatrick 2010). In electronic media there is no scarcity of space and thus no rationale for filtering before publication. Instead, an arena is created where everything can be accepted and the participants in that space either take up the offering or not, help refine it through debate and disseminate it around the community. This is filtering by the many rather than the few and examples exist of such systems working in science and engineering (the arXiv repository for instance, see Fitzpatrick 2010). We are suggesting that such an open forum precede the normal review process for future AaeE conferences. Such a forum has the potential to allow broader participation and the identification of wider interests and topics of more general interest, as well as to model and nurture better research practice. There is ample evidence (Fitzpatrick 2010 but see also <http://www.iiis2011.org/wmsci/website/default.asp?vc=27>) that such processes actually raise the level of debate as long as a few basic principles are adhered to. The forum must not be manipulable by a few gatekeepers. Participation must reach a critical mass and for most academics that means it must be built into institutional reward systems. We have to start arguing that this kind of activity be included in the impact measures of our work. And finally there has to be some quality control - but of reviewers, not submissions. This can be attained by identifying all contributors to the forum - ineffectual responses or flames will soon be weeded out by the community and the result will be more inclusive engineering education research practice.

REFERENCES

- Borrego M. (2007) Conceptual Difficulties Experienced by Trained Engineers Learning Educational Research Methods. *Journal of Engineering Education*, 96, 2, 91-102.
- Case, J.M. and Light, G. (2011) Emerging Methodologies in Engineering Education Research. *Journal of Engineering Education* 100, 1, 186-210.
- Fitzpatrick, K. (2010) Peer-to-peer Review and the Future of Scholarly Authority. *Social Epistemology* 24, 3, 161-179.
- Gibbs G & Simpson C. (2004) Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Goodstein, D., (2000), "How Science Works", *U.S. Federal Judiciary Reference Manual on Evidence*, pp. 66-72.
- Journal of Engineering Education* 100, 1(2011).

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 Epistemology*, 24 (3): 149-159.
SPARK^{PLUS} <http://spark.uts.edu.au>

Copyright statement

Copyright © 2011 Willey, K., Jolly, L., Tibbits, G., Gardner, A.: The authors assign to AaeE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to AaeE to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM or USB, and in printed form within the AaeE 2011 conference proceedings. Any other usage is prohibited without the express permission of the authors.