MERJ: Reviewing the Reviews

Beth Southwell
University of Western Sydney
<b.southwell@uws.edu.au>

The Mathematics Education Research Journal is a highly regarded international research journal that allows contributors to disseminate their research findings to a wide audience. The peer review process that contributes to its quality depends very much on the reviewers who for the most part are from the ranks of MERGA membership. The reviewing process is sometimes questioned by contributors so an attempt to clarify some of the critical issues involved in the process is timely. The purposes, criticisms and the criteria used are explored in relation to twenty papers reviewed during the years 2000-2003.

Reviewing articles submitted for publication in either a journal or conference proceedings is a very important task. The discussion that has taken place in recent times is more about the process than the actual purpose of the process which is not always considered the same by those involved. In general, most would see it as a means of quality assurance, though others might see it as a way of getting assistance to improve their work. The latter are, I would conjecture, in the minority. Other views expressed by Dalton (1995) include metaphors of policeman, hatchet man, prosecuting attorney, defence attorney or even the Grand Inquisitor. Dalton cites Manheim (1975) who suggests that the purpose of the reviewing process might be to "take the heat off editors by absorbing the responsibility for unfavourable decisions (p. 219)."

Scholars contribute to research journals for different reasons also. The main reason seems to be the dissemination of their research findings and in that way, establishing a precedent in the form of new knowledge, and putting new knowledge into the public domain for critique and development. Another reason is in terms of personal gain for the benefit of enhancing a curriculum vitae or other advantages. (Gorman & Breen, 1999; Gorman and Calvert, 2000).

Some researchers do not demonstrate good communication skills and are sometimes more interested in ideas than writing about them. When constrained for whatever reason, the reports do not always reach acceptable standards of expression and style.

Most editors of scholarly journals that hope to provide quality literature for their readers have a definite idea of what is to be expected in such quality contributions to the literature. Many academics, however, have known the sometimes unpleasant experience of having their paper rejected by one reviewer while it is accepted and even praised by another reviewer. Many will also have had the experience of being a reviewer and having difficulty for various reasons in deciding whether to recommend rejection or not. Having been involved in several instances where there has been anomalies of these kinds, the author decided to investigate the basis upon which the contributions to the *Mathematics Education Research Journal* (MERJ) are judged and how it compares with criteria used by other editors and journals.

Misconceptions and criticisms

Several misconceptions have arisen concerning the reviewing process for journal articles. These include criticisms concerning the objectivity of the reviewers, and their motivation in agreeing to be reviewers. Another criticism is that refereeing is unreliable

since the reviewers do not agree with each other, and is inefficient because it delays publication. Some of these criticisms were thought to be overcome by the process of peer or blind review but there have been several criticisms of that process also. Justice, Berlin, Fletcher, and Fletcher (1994) questioned the match between the readers' and the reviewers' judgements of an article and found that a discrepancy did exist. Kassirer and Campion (1994) maintained that the process of peer review was not perfect because it does not eliminate bias, either by the editor or the reviewer, but is, nevertheless, indispensable. This view is supported by Ryan and Martinson (1999) who recommended that editors should establish an effective and efficient review system and included a list of processes they see as essential for this. The list is:

> Acknowledge receipt of manuscripts and evaluate them promptly, respond promptly to information requests, refuse to blame referees for delays, refuse to send referees manuscripts that the editors consider unacceptable, state clearly why manuscripts are rejected, specify how accepted manuscripts might be improved, and protect authors against poor reviewers. (Ryan & Martinson, 1999, p. 2)

Feurer, Becker, Picus, and Ramirez (1994) measured the reliability and preliminary validity of a grading instrument for editors to assess the quality of peer reviews on the assumption that such a grading instrument would be useful to editors in evaluating the recommendations of reviewers. They developed the instrument and established the reliability and validity statistically.

Ethics and Reviewing

One of the major elements that impinges on the reviewing process and indeed on educational research in general is that of ethics. The Australian Association for Research in Education (AARE) has established a Code of Ethics for research and includes a section on "Editing, Reviewing and Appraising" (p. 125). The emphasis is on standards of scholarship, objective judgements, constructive and educative reviews, and maintaining of ethical standards. Universities all have Human Ethics Protocols but these do not usually include statements about the reviewing of academic research. Dalton (1995) raises ethical issues such as possible conflict of interest between author and reviewer, competition, possible plagiarism, the need to adjust for reviewer bias, and the need for training of reviewers.

Criteria used by MERGA and PME

The criteria on which contributions to MERJ are judged are straightforward. The form that reviewers receive carries a list of criteria as presented in Figure 1.

MERJ manuscript number:
Title:
Review requested by _____
I rate the manuscript as follows (delete all but one):
Recommended for publication subject to minor changes to be made in consultation with the Editor.
Recommended for publication subject to revisions to be carried out to the satisfaction of the Editor.
Recommended for publication only after the author has rewritten the manuscript and the paper been reviewed again;
Not recommended for publication.
My reasons, taking the manuscript as a whole:
Significance:
Distinctiveness:
Relevance:
Methodology:
Length:
Organisation:
Writing:
Style:

- Specific weaknesses, with suggestions for revisions:
- How the manuscript might be rewritten (if major revision recommended):

Figure 1. Criteria for reviewing contributions to MERJ.

As well as these criteria, reviewers are provided with guidelines which set out the types of comments that would be helpful.

Reviewers of papers for presentation at PME conferences and publication in the book of proceedings are straightforward as set out in Figure 2.

Proposal Title:
Please mark your recommendation below:
Check this box [x] if you would like to see other reviews of this paper, if at all possible, recognising the administration time and expense.
Recommendation: [X] ACCEPT [ ] REJECT
Comments:
Theoretical framework and related literature
The methodology (if appropriate)
The statements and discussion of results.
Clarity
Relevance to the PME audience
Detailed reasons for recommendations and general comments.

Figure 2. Criteria for review of conference papers for PME.

PME also have a set of criteria for reviewing theoretical or philosophical essays as distinct from empirical research studies. These are given in Figure 3.

Proposal Title:
Check this box[x] if you would like to see other reviews of this paper, if that is at all possible, recognising the administration time and expense.
Recommendation: [X] ACCEPT [ ] REJECT
Comments

The argument of the paper and its rationale:
Review of literature (including analysis and critique of frameworks informing the argument):
Soundness of the argument:
Clarity:
Relevance to PME audience:
Detailed reasons for recommendation and general comments (e.g. suggestions for presentation if recommended for acceptance or clear reasons for rejection if recommended for rejection):
SO or PP. If appropriate, give reasons for suggesting the paper as short oral communication [15 minutes] or a poster presentation:
Research categories. If you think the categories given by the author (s) are not the most appropriate, give two alternative (numbers) below.
Other comments.

Figure 3. Criteria for review of theoretical or philosophical conference papers for PME.

The reviewers for PME are also sent some examples of helpful and unhelpful reviews for both recommendations for acceptance and for rejection.

General Literature

The general literature on mathematics educational research carries little in terms of information on reviewing of journals. In recent years, the only relevant Australian paper on a similar area is that written by Truran and Truran (2000). They examined the extent to which reviewers were fair to the contributors of papers to the 1999 MERGA Conference. They agreed that it had been a fair process but advocated some possible improvements to the process. These included detailed planning and more editorial authority. Dunkin (1996) reported a study of errors that had been made by synthesisers of research in the area of professional growth. He found that nine types of error occurred at different points in the reviewing process. They ranged from such errors as wrongly reporting sample size to stating unwarranted conclusions from the research reviewed. The most prolific sources of review criteria apart from these appear to be the *American Sociologist* and the *Journal of the American Medical Association* (Dalton, 1995). Another source is the *International Federation of Library Associations and Institutions* (IFLAI).

It is recognised that the reviewing process for conference papers is different to the process for reviewing journal articles. For one vital thing, journal papers are usually 3 or 4 times as long and this gives the writer the opportunity to develop arguments much more fully and more forcefully. Because of these differences, the writer of this paper has decided to concentrate on journal articles only, at the same time using the criteria specified for conferences in order to identify common assessment criteria.

Comparison of Review Criteria

There are certain elements in the above Figures 1-3 that indicate a commonality of expectation in the reviewing processes for MERJ and PME. These appear to be that the article being considered include the following:

- Some form of rationale in the form of a theoretical framework, or statement of significance and relevance;
- review of literature;
- some form of methodology;
- development of an argument or analysis of results;
- clarity; and
- satisfactory style.

There are some journals that do not set any particular criteria and rely completely on the judgement of the reviewers to make recommendations in the light of their experience and expertise. In fact, it has been known for some reviewers of articles for MERJ to ignore the set criteria and develop a statement according to their own insights. As the set criteria form a guide, more experienced reviewers find them too restrictive and prefer to write comments in their own style.

Gorman and Breen (1999) and Gorman and Calvert (2000) list six different criteria for assessing submissions to a journal with only slight variations in order in the two lists. The six criteria in the 2000 list are:

- new information or data
- level of scholarship
- advancement of knowledge
- acceptable research design
- theoretical soundness
- appropriate methodology and analysis

There are a number of similarities in these two lists. *Level of scholarship* could cover a range of ideas, including the review of literature and clarity. No doubt different researchers and editors will want to express their understanding in different ways.

Analysis of Reviews of MERJ Contributions

For the purposes of this paper, only papers reviewed in the years 2000-2003 are being considered. Also, it needs to be noted that these comments refer to only approximately one third of the papers submitted to the MERJ editor for consideration and also ones within a limited range of topics.

Types of papers that are acceptable cover a wide range of research activity. The instructions in MERJ itself ask for "high quality papers in any areas of mathematics education research" (inside back cover). Another way of categorising contributions is according to the topic and that has been attempted also in Table 1 using as much as possible the categories listed for the MERGA Conference.

The data given in Table 1 indicate the range of topics, though it must be stated that types listed are not mutually exclusive and, in general, cross types.

Table 1.
Categories of papers submitted

Category	*Recommendation*			
	A	B	C	D
Beliefs and attitudes	1		2	
Learning strategies		1		
Cross-cultural studies		1		
Tertiary education		1		
Applications and modelling		1	2	
Problem solving		1	1	
Advanced mathematical thinking			2	
Number concepts			2	
Curriculum			1	
Values			1	
Technology/teaching style				1
Geometry				1
Statistics				1
Totals	1	5	11	3

Table 2 sets out a breakdown of the number of papers received, whether they were rejected or accepted and the ratings finally given.

Table 2
Analysis of Submissions to MERJ for 2000-2003

No. submitted	Recommendations				No. published	No. overseas reviewers	No. reviewers replaced
	A	B	C	D			
20	1	5	11	3	6	6	7
	5%	25%	55%	15%			

The number of reviewers that had to be replaced is high (approx 12%). The replacements were necessary for various reasons. In some cases, the reviewer had moved institutions and had no electronic forwarding address. In other cases, the reviewer was on study leave unknown to the Associate Editor and did not respond at all. After a reasonable amount of time had elapsed, a replacement reviewer needed to be obtained. In some few cases, the intended reviewer contacted the Associate Editor, as requested in the guidelines reviewers receive, to say they would not be able to complete the review in the time specified. In general the overseas reviewers were prompt in returning reviews except in one case where an automated email message alerted the Associate Editor to her being on leave. It is obvious from Table 2 that the majority of recommendations and subsequent final outcomes was in the C category. This is the category that would require the author to do most work to have it reviewed again and it is from that category that the greatest fallout occurred. So far, only six papers have been published from the 20 reviewed. These six comprise four that resulted in A and B outcomes and only two of the C outcome.

An analysis of all the reviewers' comments for each of the 20 papers reviewed revealed that there were nine distinct types of comments made by reviewers in their evaluation of the papers. These were then allocated to each of the ratings B, C or D recommended by the reviewers. These data are shown in Table 3.

From Table 3, it can be seen that the criticism that occurs most frequently relates to the methodology. Comments mainly include insufficient detail, and lack of clarity. The criticisms related to a theoretical framework are mostly that it is non-existent. Literature background is often too brief, not described fully or not linked with the rest of the article. The number of comments relating to English and style are relatively few, especially considering the number of papers from non-English speakers.

Table 3
Distribution of referees' comments according to assessment criteria and recommended rating

Type of comment	Recommend B	Recommend C	Recommend D	Total
Framework/rationale	2	5	3	10
Literature support	3	5	1	9
Methodology	1	10	2	13
Analysis	2	4	1	7
Discussion	2	2	2	6
Clarity	1	4	1	6
Definitions	1	4	0	5
English/style/length	0	2	2	4
Unsubstantiated statements	0	4	1	5

A further analysis of the results of the reviewing process was in terms of the outcomes for each year. These data are shown in Table 4.

Table 4
Distribution of review outcomes for the years 2000-2003

Year	Final Result				Total for year
	A	B	C	D	
2000		3	4		7
2001		2	2	1	5
2002	1		2	1	4
2003			3	1	4

On the other end of the scale there were four papers on which the reviewers reached consensus. One outcome was an A, one a B and the other 2 were C. This is a yield of 20% consensus on the papers reviewed and that is considered quite satisfactory. There was a major variance between reviewers in the case of only four papers.

The number of papers submitted seems to have decreased slightly over the four years. This could be due in part to the number of special theme issues that have been edited in the same period of time.

Conclusions and recommendations

The reviewing process for articles for MERJ may have its problems but it is basically sound. This analysis, however, indicates some recommendations for contributors to research journals to observe. The criticisms made by a range of reviewers indicate that contributors need to pay more attention to the detail in their methodology and ensure that a theoretical framework or rationale is well enunciated.

Contributors also need to adopt the mental attitude that rejection of a paper at whatever level is the opportunity to develop the paper further and not to abandon it. The recommendation concerning the training of reviewers should not be ignored and could be conducted online for a brief period of time. Alternatively, a training session at the Annual Conference is a possibility. The skill of the reviewers will, in the end, determine the quality of the journal.

References

Dalton, M. S. (1995). Refereeing of scholarly works for primary publishing. *Annual Review of Information Science and Technology (ARST), 30*, 213-250.

Dunkin, M. (1996) Types of errors in synthesising research in education. *Review of Educational Research*, *66*(2), 87-97.

Feurer, I. D., Becker, G. J., Picus, D., & Ramirez, E. et al. (1994). Evaluating peer reviews: Pilot testing of grading instrument. *Journal of the American Medical Association, 272(2), 98-102.* http://www.proquest.umi.com/pqdlink?Ver=1. Retrieved 14 March 2003.

Gorman, G. E., & Breen, E. (1999). *Library and information science journals in the Asian contexts*. Paper presented at the 65th International Federation of Library Associations and Institutions Annual Conference, Bangkok, August, 1999.

Gorman, G. E. & Calvert, P. J. (2000). *Journal quality in the Asian region: Results of a pilot study for the IFLA Round Table of library and information science journals*. Paper presented at the 66th International Federation of Library Associations and Institutions Annual Conference, Jerusalem, August, 2000.

Justice, A. C., Berlin, J. A., Fletcher, S. W., Fletcher, R. H., & Goodman, S. N. (1994). *Journal of the American Medical Association*, *272*(2), 117-120. http://www.proquest.umi.com/pqdlink?Ver=1 Retrieved 14 March 2003.

Kassirer, J. P., & Campion, E. W. (1994). Peer review: Crude and understudied, but indispensable. *Journal of the American Medical Association, 272*(2), 96-98. http://www.proquest.umi.com/pqdlink?Ver=1 Retrieved 14 March 2003.

Ryan, M., & Martinson, D.L.(1999). Perceived problems in evaluation of mass communication scholarship. *Journalism and Mass Communication Educator*, *54*(1), 69-75. http://www.proquest.umi.com/pqdlink?Ver=1 Retrieved 14 March 2003.

Schulman, K., Sulmasy, D. P. & Roney, D. (1994). Ethics, economics, and the publication policies of major medical journals. *Journal of the American Medical Association*, *272*(2), 154-158. http://www.proquest.umi.com/pqdlink?Ver=1 Retrieved 14 March 2003.

Truran, J. & Truran, K. (2000). Is the MERGA Conference refereeing process fair? In J. Bana & A. Chapman (Eds.). *Mathematics Education Beyond 2000* (Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia, Fremantle, pp. 596-602). Sydney: MERGA.