Is the MERGA Refereeing Process Improving the Quality of Australasian Mathematics Education Research?

Kathleen Truran
University of South Australia
<Kath.Truran@unisa.edu.au>

John Truran
Private Practice, Adelaide
<truranjk@camtech.net.au>

This paper follows from a belief that the activities of an educational organisation should model good educational practice. MERGA's refereeing practice is examined as a way of improving the quality of teaching and learning about mathematics education research for both authors and referees. While many positive features are already present, some weaknesses or omissions are identified which, it is suggested, are more likely to be addressed if MERGA's educational role is seen as an important focus of its work.

In the calls for contributions to annual MERGA Conferences will be found statements like the following:

> MERGA believes that the review of papers gives conference attendees some confidence in the quality of the publication of proceedings. The refereeing process also provides a mechanism for peer review and critique and so contributes to the overall quality of mathematics education research.

This objective matches the first official aim of the Group: "promoting quality research in mathematics education" (MERGA Constitution, 10 Jul 1997). MERGA, as a professional organisation, wishes to ensure that papers presented and published at its conferences conform to high academic standards. But how are such standards and quality to be achieved? Osmosis is rarely sufficient. We will argue in this paper that when MERGA's activities, in particular its Conference refereeing process, are seen as providing models of good educational practice, then the question being addressed here becomes an important preliminary to identifying ways of enhancing MERGA's current practices.

The review process used is described in J. & K. Truran (2000); here we assess its value at the 1999 Conference for improving authors' papers, especially in the case of Early Bird (EB) submissions where resubmission is possible. This is clearly a teaching process, one conducted at a distance and, quite unusually, one where teacher and learner are unknown to each other, in theory at least. It is an area which has attracted little formal research attention, partly, of course, because it is conducted confidentially. This paper uses confidential data available to us as editors of the 1999 Conference Proceedings (J. & K. Truran, 1999) to examine aspects of this rather special teaching/learning situation. We report in a way which illustrates current practices while preserving anonymity.

The Refereeing Process

Three referees competent in the field review each paper independently and all reviews are then sent to the author to support the editors' acceptance or rejection of the paper. While the review relies on subjective judgements, reviewers are asked to assess the paper against seven criteria, using a broad interpretation of the accepted norms for reporting research, and ensuring that it represents a contribution to mathematics education or research processes.

The success and perceived quality of the papers presented depends on the quality of the referees and their knowledge of the subject that they review. This is an issue attracting some attention in mathematics education research circles. For example, the Chair of the 20th annual Psychology of Mathematics Education Conference reported that some authors of rejected papers at PME conferences were claiming that some of the accepted papers were of lower quality than their own (Gutiérrez, 1998). Our subjective impression after attending five of the

seven MERGA refereed Conferences is that similar views about poor papers being accepted are also found in MERGA, and not only by authors of rejected papers.

Gutiérrez also said that PME, which also relies on members volunteering to referee for its conferences, had found that some referees seem to confuse *being* expert in a topic with a *wish to be* expert in the topic (italics added). He stressed that in his opinion reviewing papers for PME was not the way to acquire expertise. We have suggested in J. & K. Truran (2000) that a similar confusion may exist among some MERGA members.

Nevertheless, none of us has finished learning about our research interests, and all referees have incomplete knowledge, even about their areas of specialisation. So, while the quality of referee's knowledge and comments is basic to the success of any conference and its publication, "quality" cannot be seen as a dichotomous attribute, but rather as a developing one. This view fits comfortably with one which sees the refereeing process as a broad educational activity.

In this paper we first examine the EB Process which provides an opportunity to see how authors use referees' comments to change their papers. We then look at some of the reviews presented for the standard papers to identify the types of comments made and assess their value for the authors, and we conclude by looking at what seem to us to be some missed opportunities for further educational development.

The Early Bird Process

EB reviews are available to inexperienced authors who feel the need for an opportunity for rewriting if judged necessary. In 1999 several experienced authors submitted EB papers as well: this may indicate some lack of confidence in the review process, but we have no specific evidence for this, and know that in two cases it was related to the planning of Australian visits from overseas.

Of the seventeen EB papers submitted, eight were accepted at once, and nine were returned for some revision. It was relatively easy to match EB papers with appropriate referees, because these papers were the first to be allocated. Referees recommending revision were asked to provide more detailed comments than they might for more experienced authors. Most referees also agreed to review the revised papers, which made for more consistent reassessment.

Examples of Referees' Comments on Early Bird Papers

We have probably all had to cope with submitting a paper that has been worked on and sweated over for a long time and receiving a "not accepted" judgement. Sometimes the referees' comments cause irritation and disbelief—how could someone have said what they did; how could they have missed the paper's real intent? Such irritation is exacerbated when the review is less than objective or indicates any weaknesses in a referee's knowledge. While MERGA provides a pro forma for use of referees, not all use it, and not all address all the criteria. The loss of educational potential in such cases is obvious.

For the 1999 EB papers there were examples of both exemplary and less than useful reviews. In an extreme case just one word—"accept"—was all that was given. Interestingly, some short acceptances were given to papers where careful responses from other reviews required a subsequent re-write of the paper. But in other cases all three referees took considerable trouble, providing useful, supportive comments presenting the author with examples of common flaws, omissions or illogicalities, and indicating a useful base for modification and eventual acceptance of the paper.

Comments like "the main reason for asking you to rewrite the paper is to do with length and layout; but as you have the chance to revise there are a couple of minor points that you

could fix as well ... " show a concern for the author and also an understanding of the subject matter. This referee made suggestions about every aspect of the paper, pointing out superfluous comments made by the writer as well as the need for further discussion on some points.

As well as providing detailed comments within the pro forma provided, extensive comments were sometimes written on the actual papers. This strategy is probably of even more value for inexperienced authors. Comments like,

- a statement is required here that summarises the apparent benefits,
- this is the important one the others could probably be omitted,
- I suggest that these be reworded, and
- not sure how this follows or what the point is, perhaps clarify or omit;

directed the attention of an author to a specific point in the paper. Some referees checked mathematical calculations and were able to encourage authors to consider re-checking these before resubmitting. For example, "I carried out the usual *t*-test method and found slight disagreement with the ranges given." Comments like these would have helped the author not only to rewrite the paper but also to polish it so that it became a very good paper indeed. It was evident from re-submitted EB papers that most authors took the majority of the comments and suggestions seriously.

Of course, other authors had less support, particularly when referees merely ticked each of the criteria sections or provided imprecise comments like "some sentences are quite long and maybe shortening them would be valuable". Fortunately, such cases were a definite minority.

Changes Made on the Basis of the Referees' Comments

Sometimes the revisions required were technical and easy to implement. But in others referees gave detailed comments such as "the methodology could be clearer. ... An example of a question/problem would help here. ... I think that the text about each student [could be included in a table]. ... I suggest that [the following] be reworded." (Quotations have been slightly modified to preserve anonymity.) Authors took most of these criticisms to heart, and we were amazed in some cases just how different the original and revised papers were. As one referee observed, "The author has acted on comments in my review and the revised version is now dramatically improved." Unfortunately, the requirement of anonymity prevents us from citing particular examples.

While there were examples of changes *possibly* made to please the referee and ensure acceptance, there were other examples of creative rethinking in response to criticism which is at the heart of the educational process. For example, the suggestion quoted above to incorporate a large amount of text into a table was only partially followed; other relevant details were introduced into the narrative at appropriate places. In other words the referees' comments were taken as observations about weaknesses which needed to be corrected, but the author decided his or her preferred way of correcting the weakness.

A particularly interesting case where authors did not always take up referees' comments occurred when referees pointed out the limitation or omissions in the References. Sometimes they suggested that "a wider reading list should have been included", but gave no suggestions about appropriate references. Such a comment could easily be beyond the knowledge, experience or paradigm of the author and very difficult to act upon. In other cases, "missing" references which were specified by the referee were still not included in the revised papers. This may have been because of the time required to find and incorporate new ideas or perhaps a difficulty in understanding why the referee made the suggestion. Turn round time for resubmission was tight and there is no opportunity for discussion with the referee. So

frequently the resubmission contained unchanged References. Authors who decide to do this should realise that some referees do suggest that their own work should have been cited!

While we have focused on the authors' reactions to thoughtful reviews, we were also sent comments by a small number of "first time" referees saying how much their own reviewing skills had been improved as a result of having been involved in the EB process. They appreciated the opportunity to read a paper carefully, consider and write their own review and also see the modified paper.

So we may say that the EB refereeing process led to significant improvement in several cases. At its best it was of educational value for both authors and referees. But a few referees did not see their work as contributing to a rich educational activity, and even the best reviewers made some impracticable suggestions given the time constraints involved. While there is room for improvement, and especially for more reflection time, the EB process as it stands may certainly be seen as an effective educational model for refining skills.

Special Features of Reviews of Standard Papers

Three differences from EB reviews may be seen in those of standard papers. Putting aside minimalist reviews clearly done under time stress, rather more standard reviews were relatively brief, especially if they were positive judgements. Some of these came from referees who had prepared very detailed reviews of EB papers. One referee actually wrote, "because I recommend that this paper be accepted my comments are short." Referees may not be aware of how often editors must consider and weigh up quite conflicting reports. For example, three referees of one paper responded in totally different ways:

- this paper is well written and addresses an area of concern;
- very unimaginative, purely statistical analysis;
- why was this topic even investigated?

Editors do need to read reviews to ensure that no serious injustice is being done. This is not easy. Another paper received a damming negative review from one referee, with another reviewer making *the same observations* but with an "accept" decision. Carefully argued cases for positive decisions are as valuable both to editors and authors as those for negative ones.

The second difference we noticed was that many referees of standard papers tended to give authors the benefit of the doubt. EB referees could recommend revision, but standard paper referees could not, and were well aware that failure to accept a paper could have significant implications for an author's funding and attendance. While many were prepared not to accept a paper, others, perhaps those not really expert in a particular field, probably took the editors' advice not to require journal article standards and decided in an author's favour. For example, two positive reviews of one paper wrote

> The standard of English is not very good. On the other hand the academic content of this paper warrants some dissemination and discussion; so I lean towards acceptance.

> This is an interesting paper with a constructive purpose and some helpful results. ... It would benefit from a rather tighter and more general conclusion, and from being set within a wider base, but it represents a useful step in the direction of making research relevant to harassed classroom teachers.

The absoluteness of a negative decision probably also influenced responses to weaknesses of layout. In one case all three referees indicated many typographical and grammatical errors and unclear segments in the paper, but all three recommended acceptance. But in other cases, similar errors were the only basis for a negative decision. It is possible, however, that some authors may be presuming on this prejudice in their favour. One referee wrote:

> There needs to be some insistence that writers have their work read by another before they send it in. Noting poor grammar and spelling should not be the responsibility of reviewers. It does not happen to the same extent with journal articles that I review, so why with Conference papers. Are they not meant to be of similar quality?

Papers are often prepared under time pressure, so technical issues are likely to receive minimal attention. Not only does there need to be more consistency about judging such errors, but, as we shall see below, inadequate attention to such details may prejudice an author's chances of acceptance.

The Value of Referees' Comments for Standard Papers

It is difficult to assess the value of the refereeing process for standard papers, because little feedback is available and the diversity of responses must be confusing for some authors. But some general comments can be made.

Tables

Tables are a minefield for authors and referees alike. Creation, organisation and presentation of clear and accurate tables is a more difficult task that most of us would assume. Many referees mentioned accuracy, clarity, and shading of sections in tables, or the use of a table when text would have been just as effective. Comments to justify referee' problems of interpretation were often particularly clear and can be of both special and general value to authors:

- check Table 1—difficult to read;
- the two columns for this aspect should be omitted;
- the data in Table 1 can be stated more easily in the text following it;
- I recommend that ... sample sizes be indicated in Table 1;
- in the right hand column of Table 1 it would be clearer to head it ... .

Value for Authors

Just as for EB papers, some reviews were written in such detail that they would have provided authors with material for thinking about their presentation at the conference and possibly even reviewing their written presentation of conference papers in the future. As one referee wrote:

> I have been fairly critical of the paper on the assumption that it is part of a large project and probably for higher-degree studies. The faults have been identified for the benefit of the author's future work.

One referee included a personal reflection at the conclusion to a review, which may have helped the author to relate better to it. On the other hand, some reviews became so detailed, that, while they might have been useful for journal submissions, as conference reviews they were rather precious.

But we are unable to assess the extent to which authors benefited from their referees' reports. Questionnaire responses were usually happy about reviews received, but we did not ask about the influence of the reviews on authors' professional skills. A longitudinal study might be of value here.

Value for Referees

We have already noted how some first-time referees valued the refereeing of EB papers. It is very difficult to surmise the benefits in general, especially for experienced researchers. In some cases quotes made by the author have been challenged, the source of these quotes located and discussed, and the intention of the original author argued. We may presume that

such activity was of some value to the referee as well. But perhaps even the writing of a detailed review by an expert who can draw quickly on relevant knowledge may provide a valuable learning experience, especially if the paper is approaching old ideas in a new way. Educational theory would certainly suggest so. However, we are unable to provide hard data on this issue.

Some organisations distribute all three reviews to all three referees at the same time as the reviews and editors' decisions are sent to authors. In our personal experience this can be a valuable learning process, and we recommended its introduction to the MERGA executive.

Did Referees Understand the "Big Picture"?

Rejected authors frequently reported that referees were "not able to see the 'big picture' in my paper". Compressing complex ideas into eight pages is not easy. New or different approaches are particularly difficult to compress. If these ideas are very new a referee may consider the author does not understand the field sufficiently. This important claim needs to be considered in detail with reference to two such comments. The first was made by a senior researcher:

> I feel however that two of the reviewers, even the positive one, did not understand the point of my paper. Of course, that is probably my fault for not being more explicit but I had expected reviewers to be able to identify the main ideas in the papers they were reviewing. It could be they were looking at a straight qualitative or quantitative research study, which I do not believe is the only kind permissible at MERGA.

The second was made by a less experienced author who submitted a joint paper with a supervisor:

> However, suggestions such as, to include primary material as well as secondary material, and that focusing on one content area is too narrow, makes me wonder if the push [is] to go broad instead of looking more deeply at issues in Maths Ed.

Both sets of authors used quite complex theoretical structures. In both cases a small number of detailed case studies were used to illustrate the validity of the structures. The problem is obvious: detailed reports and detailed theoretical hypothesising do not fit easily together into an eight-page paper. But authors need published papers to get to conferences, and if their work is new or complex, this should not hamper their chances of acceptance.

But it is not surprising if the second author felt not to have been understood when comments were made like "perhaps some of this information could have been presented in tabular form for conciseness and to maintain the interest of the reader" and "the discussion fails to link the data descriptions to the theoretical perspective ... or to extend this perspective as well as it could". On the other hand, in our view, neither of these comments demanded more than might be achieved within eight pages using the same data and theory used by the author. The judgements were about balance, which is a reasonable thing for a referee to do. But the author's view was that "some suggestions indicated strong preferences to other structures and ideas that could not be covered in 8 pages". When different perspectives like these are presented via the anonymity of the refereeing process, authors are unlikely to change and develop as a result of referees' comments. Given that the author was happy to accept other criticisms from these referees, and that their reviews seem to us to be reasonable, there does seem to be a need for face to face discussion.

The first case is more complicated. It is true that only one referee gave evidence of being well read in the literature of the field, but both the other referees made comments about technical weaknesses in the presentation which any experienced reviewer might have made. One of these also commented

> I have some concerns about the paper, but in the end I have been swayed by a belief that this topic is worthy of a Conference presentation. A topic such as this is difficult to address purely in text form, and the discussion format of a presentation should provide an additional avenue for the exchange of ideas.

Here we can see the "benefit of the doubt" syndrome again. Nevertheless, in our view, while the author's claim that the ideas were not well understood by the referees may have some validity, the rejection was based much more on technical issues rather than on the difficulty of the topic or the referees' inability to see the big picture.

A more detailed investigation than this is needed to be sure whether referees are inclined or not to see the big picture. The evidence presented here in detail neither confirms nor refutes the claim, but does show, as we have argued above, that technical weaknesses in paper construction may divert referees' attention from the big picture to what they see as more pressing issues for a conference presentation.

Discussion

This paper has shown that the MERGA refereeing process clearly fulfils some criteria for being an effective teaching/learning process. The vast majority of the reviews are constructive assessments of the authors' work, and many provide useful advice for further development of the ideas. These reviews address both the process of doing research and also of reporting it. But teaching is much more than telling, and we have also been able to show that for the EB papers authors have been able to implement recommendations to the obvious betterment of their papers.

Such results might be seen as obvious, even trivial. But in the process of examining them deeper questions have also been raised and not always answered. These may be listed as:

- the relationship between the quality of conference papers and journal articles, especially given their perceived equality by funding bodies;
- the need for referees to give authors confidence in their expertise;
- the potential value of the refereeing process for referees;
- the importance of good presentation, especially of tables, for ensuring acceptance;
- the possible value of allowing authors and referees to make contact with each other at some stage in or after the assessment process.

MERGA is an educational organisation, so its practices should conform with high educational principles. The refereeing process is a rich environment for teaching and learning by both authors and referees, and we have shown that it is often used constructively. But we have also shown that some pedagogical opportunities are being lost which could easily be taken up. This reflects a point of view which sees MERGA not merely as an organisation to *examine* education, but puts more emphasis than at present on *practising* it. MERGA is already doing many things well to assist the life-long educational development of its members; when this is seen as an integral focus of its work, then some of the harder educational issues are more likely to be addressed, for the general benefit of all members and mathematics education in general.

References

Gutiérrez, A. (1998). Reviewing the reviewing process. *PME Newsletter*, May 1998, 10–11.

Truran, J.M., & Truran, K.M. (Eds.) (1999). *Making the Difference* (Proceedings of the 22nd annual conference of The Mathematics Education Research Group of Australasia Incorporated, held at Adelaide, South Australia, 4–7 July, 1999). Sydney: MERGA

Truran, J.M., & Truran, K.M. (2000). *Is the MERGA refereeing process fair?* Paper to be presented at the 23rd Annual Conference of The Mathematics Education Research Group of Australasia Incorporated, held at Perth, Western Australia, 5–9 July, 2000.