To Accept or Reject: Peer Review

William T. Fagan
University of Alberta

Peer review is an established and perhaps essential part of the publishing process. In this study, 1214 review comments by 216 reviewers were analyzed to determine the nature of the reviews and to get a sense of the reviewers by developing profiles based on the nature of the comments. The reviews were categorized according to tone (complimentary, constructive, or negative); direction (specific suggestions for change versus statements of criticism); and focus (the particular aspect of the manuscript which the reviewer selected for comment). In general, the topic/content and methodology were most likely to be selected for comment, which was usually negative. Whereas authors of accepted manuscripts were often given specific suggestions for change, there was little encouragement for the authors of rejected manuscripts. Four reviewer profiles were delineated: critic, conciliator, competitor, and procrastinator. As might be expected, the reviewer profiles were interrelated with the nature of the reviewer comments. Implications for authors, editors, and reviewers are suggested.

L'évaluation par les pairs est une pratique bien établie et peut-être essentielle du processus de publication. Cet article analyse 1214 évaluations faites par 216 évaluateurs, dans le but de tracer un profil des évaluateurs à partir de la nature de leurs commentaires. Les évaluations ont été classifiées sous trois rubriques: le ton: flatteur, constructif, négatif; la substance: suggestions spécifiques de corrections ou critiques; l'objectif: un aspect spécifique du manuscrit ayant retenu l'attention de l'examinateur. En général, dans le cas d'une évaluation négative, c'est le contenu ou la méthode que l'examinateur choisit de commenter. Quand le manuscrit est accepté, l'auteur reçoit des suggestions spécifiques en vue de la correction; quand, au contraire, il est rejeté, l'auteur ne reçoit que peu d'encouragement. A partir des évaluations, on peut tirer quatre profils d'examinateurs, soit le critique, le conciliateur, le compétiteur, le temporisateur. Cette analyse pourra éclairer le travail des auteurs, des éditeurs et des évaluateurs.

When authors submit manuscripts to refereed journals, they know that the decision to accept or reject will be based on the process of peer review. The purpose of this study was to investigate reviewer comments, and on the basis of such comments, to develop a "sense" of reviewer.

Researchers generally agree that peer review is an essential practice in publishing research findings. Commoner (1978) states that peer review is more than a housekeeping task. It is necessary as a check on faulty data and unsound interpretation. Peer review has also acquired a definite prestige among professionals. Marsh and Bell (1981) state, "The process is so highly valued that some indices

of research productivity count as publications only articles that have been selected by peer reviews'' (p. 872).

The review process is based on the assumption that the system provides a check on quality (Wilson, 1978). However, this may not always be a valid assumption, for as Banner (1988) comments, ''the system is beset with many problems — of attitude, administration and effectiveness'' (p. 189). To this list Commoner (1978) adds referees' incompetence and the abuse of the privilege of anonymity by putting down a competitor or a disliked researcher. Consequently, Commoner argues that the peer review system may at times retard the progress of knowledge. As an example, Commoner refers to Dr. Rosalyn Yarrow, a Nobel Prize winner, whose first two papers describing the scientific work for which the prize was given were rejected. One reviewer commented on ''the dogmatic conclusions set forth which are not warranted by the data'' (p. 25).

A significant problem of peer review is the low inter-referee reliability of manuscript ratings (Inglefinger, 1974; Mahoney, 1977; March and Bell, 1981). Based on an analysis of the reviews of 1572 papers, Zukerman and Merton (1971) indicted that the frequency of agreement by reviewers was barely above that expected by chance.

Banner (1988) notes the effects of peer review, from its influence on the careers of individuals to its effect on the reputation of institutions. Commoner (1978) draws an analogy between a rejected manuscript and venereal disease. He states:

> It is a far more widespread phenomenon than one would guess from the frequency of personal accounts. Most victims are too ashamed of the event, or too worried about its effect on their careers to talk about it. And like VD, the rejection of proposed or completed scientific research by one's 'peers' is an important sociological phenomenon as well as a personal one. (p. 25)

The fundamental purpose of peer review should be the advancement of knowledge (Banner, 1988). Mahoney (1977) cautions that because of the significance of the role of peer review, tradition is not enough reason to defend it. ''We must understand the processes and effects of our conventional practices, and — more importantly — that we are negligent if we fail to study them'' (p. 174).

Method

Sample. It is difficult to obtain a representative sample of manuscript reviews. A random sample would have to be directed towards editors rather than authors, since the population of the latter would be difficult to determine, especially if manuscripts were rejected. Obtaining samples from randomly selected editors would depend on their cooperation and on the ethics guidelines which govern their release of reviews. In many of the studies cited earlier in this paper, reviews were obtained from a single editor/journal. While a large number of reviews may be collected in this manner, the papers would have been exposed to a small

number of reviewers. Thus there would be the possibility of reviewer bias if some reviewers rated several manuscripts.

For the present study, the author personally sought reviews from colleagues. As well, a request for reviews was placed in a national reading-language journal. Two hundred and sixteen reviews for 108 manuscripts (46 accepted, 62 rejected) were received from various parts of Canada and from several American states. There were two reviews each for 104 manuscripts, three reviews each for two manuscripts, and a summary of reviews by the editor for two other manuscripts. All manuscripts were research based and focussed on some aspect of reading-language-literacy. All comments were in English.

The sample contained twice as many manuscripts submitted to Canadian journals (allowing two reviews per manuscript) than to American journals or journals from other countries. This is understandable, considering that most of the manuscript authors were Canadian.

Procedures. Each review was coded according to the decision of the reviewer to accept or reject. Eighty-seven percent of the decisions to accept were qualified.

All reviews were then divided into comments or topic statements. The specific topic of a sentence was identified and all other sentences elaborating or extending this topic were considered as a single comment. Very often a single sentence constituted a comment, and in a few instances two comments were included in a single sentence. Comments are separated by slash lines in the following example.

> Clarification of the sample size would have been helpful. Why 20 subjects?/ Test reliability is not clear./ The distinction between slow versus average reader is not clear. Was this an arbitrary decision? Could you report standard errors of measurement for the means of each group?/

The percent of agreement by two independent raters for comment division was 99.4. There were a total of 1214 comments, or an average of 11 per review. Five hundred and sixty-six comments were made on manuscripts accepted for publication and 648 on rejected manuscripts.

Each comment was then analyzed in terms of tone, direction, and topic focus. Finally, the total reviews were analyzed to infer a sense of the reviewer. Interrater reliability by two independent raters for 15% of the comments ranged from 90% to 93.7% agreement.

Results

Tone of Reviewers' Comments. The majority of comments on manuscripts accepted for publication were either complimentary (38.5%) or constructive in tone (43.8%). As might be expected, these categories represented the smallest number of responses on manuscripts rejected. Complimentary comments included positive statements about the manuscript: "The title is clear and accurately reflects the content." "This manuscript presents a fascinating study of young

children as writers." Constructive comments contained suggestions for improving the manuscript: "Even though the author describes the subtests of the SHEL test, it would be much easier for the reader if the test was included in the appendix." "Possibly a more appropriate heading for Language Analysis could be 'Need for a Theory of Language' or 'Theory of Language.' This is the principal focus of this part of the article."

Negative comments made up almost two-thirds of the comments on rejected manuscripts. Examples of negative comments were: "The same lack of awareness of relevant literature is indicated in the author's comments about the results of the study." "I am sorry to say that this manuscript is in dire need of extensive editing. It is extremely difficult to follow."

Since complimentary and negative comments do not provide much direction for improving the manuscript, the constructive comments were further analyzed in terms of their direction or use for reworking a manuscript.

Direction of Reviewers' Comments. Reviewers who recommended that manuscripts be accepted were inclined to give constructive criticism in the form of specific suggestions for improving the manuscript. Over one-half of all constructive comments were of this type: "The fact that some children had attended kindergarten and some had not should be explained in light of their differential awareness of print knowledge." These reviewers were also more inclined to elaborate on their specific suggestions for change; that is, more than one sentence was given regarding the suggestion. Seventy-three percent of all elaborations on accepted manuscripts occurred in conjunction with specific suggestions.

Reviewers who rejected manuscripts were more inclined to provide comments of criticism only: "The literature review needs to be expanded." These reviewers were also more likely to elaborate on this kind of comment: "The author can't expect the reader to appreciate the significance of the study, when so few references were cited. In addition, those cited do not represent the major studies in this field."

Focus of Reviewer's Comments. Regardless of whether manuscripts were recommended for acceptance or rejection, most of the reviewers' comments were on two aspects of the manuscript: the subject matter or content and the methodology (see Table 1).

An example of a subject matter/content comment on an accepted manuscript was: "The author presents information in the area of multicultural education that should be read by all American, Canadian, and English educators, plus other language distinct groups." Comments of this type on rejected manuscripts included: "This article deals with an aspect of reading in need of research." "The problem being investigated has merit." However, most comments on the subject matter/content of rejected manuscripts questioned the significance of the topic:

Table 1
Topic Focus of Reviewer's Comments

	Accepted		Rejected	
	Total	**Percentage**	**Total**	**Percentage**
Subject matter/content	144	(27.1)	200	(30.8)
Sample	38	(7.2)	65	(10.0)
Methodology	146	(27.5)	240	(37.1)
Organization/Style	76	(14.2)	50	(7.7)
Implications	76	(14.2)	73	(11.2)
Technical	52	(9.8)	20	(3.2)
Non-topic related	34		0	

"Although the approach used was novel and of potential interest, the lack of important findings does not warrant publication." "Of merit, but I find nothing new." "This case history adds nothing new to the existing literature."

The largest number of comments on rejected manuscripts dealt with methodology. Many of the comments referred to lack of information: "While instrumentation is well detailed, data gathering procedures are so sketchy, replication could not be considered. More details are needed." In some cases the comments on methodology were directed not to the reporting of the research study but rather to the study itself: "Only one measure of reading achievement used to classify the subjects is a serious limitation." Still other comments referred to the reliability of the data: "The methodology worries me — no reference is made to reliability of the tests. This should be addressed; if none were undertaken, the paper should be rejected."

Comments of a technical nature occurred least often on both accepted and rejected manuscripts. These comments focussed on typographical errors, omission of references in the bibliography, or bibliographical style. Sometimes the reviewer played the role of editor in making changes of a technical nature. There were sometimes negative comments about the suitability of the bibliographical style for the particular journal. Thirty-four of the comments on accepted manuscripts were not topic specific: "I enjoyed reading this paper." "The author seems to have worked hard on the study."

Reviewers of rejected manuscripts were more likely to make comments regarding the population sample in a study than were reviewers of accepted manuscripts, while the latter were more inclined to comment on organization/style and implications. Comments on the sample either asked for more information or questioned the size of the sample in terms of the generalizability of the findings. Comments on organization/style ranged from negative remarks such as "Although

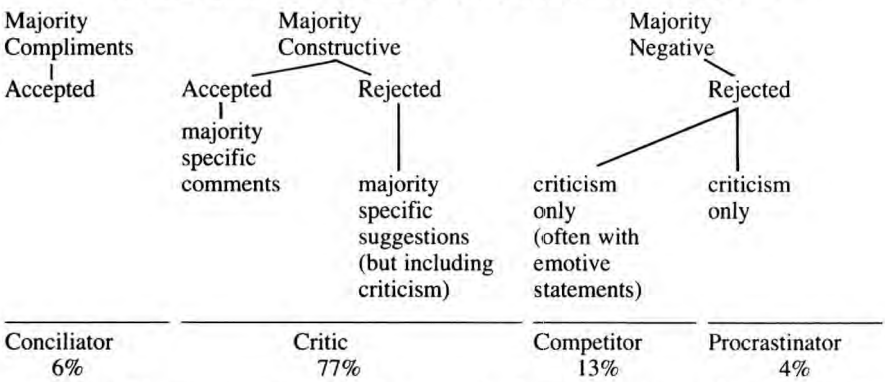
the author(s) have attempted to develop a theoretical basis for the study, the connection between the theoretical background and the methodology is not clearly and logically explained'' to more specific comments such as, ''Much of page two is written in a rather more informal style than elsewhere in the paper and sound intuitive rather than based in the literature. Could you revise or shorten this section to make it more consistent with the general style of the paper and the usual conventions of a research paper?''

Reviewer Profiles

In order to get a sense of the reviewer, each review was first coded in terms of the tone of the majority of the comments: complimentary statements, constructive statements, and negative statements. Those with a majority of complimentary statements had all recommended acceptance, while those with a majority of negative statements had all recommended rejection.

Those reviews with a majority of constructive comments were then divided according to whether the decision was to accept or reject. The ''rejection'' reviews were divided into two subgroups. Those with a majority of criticism statements but with some complimentary statements formed one subgroup, while those which included only criticism formed another category. A small group of reviews could not be classified according to the above criteria. These had all recommended rejection but the comments could not be easily coded in terms of tone or direction. The reviews were brief overall (a paragraph or less) and the comments were general rather than specific. The decision tree for delineating profiles is given in Figure 1.

Figure 1
Decision Tree Delineating Reviewer Profiles from Reviews



Four profiles were identified. One was similar across accepted and rejected manuscripts; one was unique to accepted manuscripts; and two were unique to rejected manuscripts. The profiles were labelled conciliator, critic, competitor, and procrastinator.

These labels were based on the characteristics of the reviews in each category, somewhat in the same manner as labels might be coined for factor loadings in a factor analysis.

Conciliator. The reviews of the conciliator tended to be complimentary. If constructive comments were made, they tended to take the form of specific suggestions. These reviews were relatively short compared to those of the critic — a paragraph compared to a page or more. These constituted six percent of the total number of reviews.

Critic. Seventy-seven percent of the reviews were considered the reviews of critics. A critic was defined as one who attempts to analyze a document in terms of its strengths *and* weaknesses. Comments on weaknesses usually contained constructive statements indicating a need for change or provided specific suggestions for change. The only thing which distinguished a critic reviewer of accepted and rejected manuscripts was that more criticisms occurred in the reviews of rejected manuscripts.

Competitor. The label competitor was suggested by the work of Commoner (1978). This type of review, which comprised 13% of the manuscripts, tended to be dominated by negative statements and/or constructive statements of criticism only. The reviews were mostly of comparable length to those of critics. The reviews provided a sense of the reviewer as one who regarded a manuscript as competition and seemed to have one goal — to uncover as many weaknesses as possible. The reviews were often marked by the presence of emotive words, typgraphical errors which often contradicted a point made, and inaccuracies. Examples of the use of emotive words included: "This manuscript is too carelessly written to be considered for publication in its present form." "This is a very disappointing paper." "I'm afraid I've lost patience with this manuscript . . . At the moment the report is too shapeless to be reviewed accurately." "I would have to rate this manuscript as unprofessional." "It seems as if this is a rewrite of a master's thesis by a first time writer." Comments beginning with the words, "It bothers me . . .," "It upsets me . . .," "It annoys me . . .," "In my opinion . . .," also had a negative overtone.

About one-quarter of the completed manuscripts were available and so it was possible to check reviewer comments against the actual data. One reviewer wrote, "The manuscript shows many signs of being hastily and carelessly written. (Journal name) should expert (sic) more care." In another instance a reviewer indicated that the findings were not generalizable as the sample was not randomly drawn. Yet on page six of the manuscript, the author had stated, "Two groups of 25 subjects each were randomly drawn from their respective populations." Sometimes the reviewer's statement was not clear: "There are what seem to be some serious problems with the manuscript at the level of coherence and clarity, as well. For instance, on the second page an assertion about the activities of illiterate readers somehow slides into a 'view' which is in keeping with one

attributed to Goodman (the 1970 citation of Goodman is not only dated, but ignores some important recent changes in Goodman's view, which would radically alter the author's list of three types of text information).'' (Author's note: the assertion was not inconsistent with Goodman's theory.) Finally, some reviewers suggested a different kind of analysis which was inappropriate. For example, a reviewer indicated that the author should have used a multivariate analysis of variance, when the ratio of variables to sample size of the study clearly did not warrant this type of analysis.

Procrastinator. A small number of reviews of rejected manuscripts (4%) did not clearly fit either of the above categories. The label *procrastinator* was given to these reviewers since four of the five reviews of this type were accompanied by letters from the editor in response to an author's inquiry about the disposition of a manuscript after what was considered by the author an unreasonable time delay. It is possible that the delay had occurred with the reviewer (although this is not known for certain) who, on the basis of the reminder, hastily made some general comments. All of these reviewers had recommended rejection. The reviews were short (a paragraph of less), were negative in tone, and included criticism only. Rather than addressing a particular topic, the comments were general in nature: ''The rationales for the study and the study groups are not fully developed.'' ''The paper is poorly written and very difficult to follow. I believe that to a greater extent this problem is due to terms and procedures being inadequately defined. I didn't find the appendix very helpful in defining terms.'' ''My concern is that I don't believe any strong conclusions were drawn. The discussion simply does not answer the question, 'So What?' Without exploring implications of research findings, the research itself becomes meaningless.'' One reviewer didn't feel the author had selected the appropriate journal with regard to topic, yet a check of issues of that journal for the two previous years showed six articles published on the manuscript topic.

Additional Findings. There were 17 instances when there was discrepancy about acceptance or rejection among reviewer's comments about a particular manuscript. The editor tended to accept the negative decision particularly if there was a majority of criticism type comments.

An analysis of the comments by editors showed that editors of British and Australian journals tended to take a greater role in evaluating reviewers' comment than did editors in North America. It was not uncommon for the former editors to summarize the comments of reviewers in a three to four page very supportive and constructive letter. Some, for example, indicated to the author which of the comments of the reviewers to consider in making changes. In at least one instance (Canadian journal), an editor indicated that bibliographical style would be looked after by the editors.

Discussion and Implications

Reviewers of manuscripts accepted for publication gave positive feedback and helpful suggestions for improving the manuscripts. However, the nature of the constructive comments on rejected manuscripts was generally of little help to the author in making revisions before submitting elsewhere. At the same time such reviews contained many negative comments. A logical conclusion might be that authors should accept the fact that their work is weak. This might be a reasonable conclusion if the authors were assured that the reviews were valid. The two areas that were most negatively criticized on rejected manuscripts (content significance and methodology) were shown by Mahoney (1977) and Scott (1974) to have lowest inter-referee reliability. Of the manuscripts in this study, 17 received one review recommending acceptance and one recommending rejection. All of these manuscripts were rejected. There were four additional manuscripts where the review's comments were inaccurate.

There is also a possibility of reviewer bias. This would likely be reflected in the reviews of competitors. Mahoney (1977) demonstrated through an experiment that "Reviewers were strongly biased against manuscripts which reported results contrary to their theoretical perspective" (p. 161). However, it may not be possible to eliminate all personal bias. According to Ingelfinger (1974), to do so would necessitate being supernatural. As Ingelfinger notes, "Since author and reviewer usually are by selection engaged in similar endeavors, they are almost unavoidably either competitors or teammates" (p. 686). Any source of error (bias or otherwise) in the peer review system is what bothers Mahoney (1977).

The findings of this study have implications for the author, the editor, and the reviewer.

Of most importance for authors is the information on topic focus by reviewers. Prospective authors would do well to check on the significance and pertinence of the content/subject matter they plan to write about. An examination of representative journals in one's field, as well as an awareness of "in-topics" in education, should provide guidance for authors in choosing content/subject matter. The description of the methodology was most frequently questioned, meriting almost 40% of the comments of reviewers of rejected manuscripts. However, there was often discrepancy in the comments on methodology, with one of two reviewers often requesting more information: description of instruments, validity and reliability data for standardized tests, specific questions used in interviews, etc. Perhaps one lead for the author is to check already published articles in the selected journal in terms of the degree of detail on methodology that is generally present in such articles.

Editors have a responsibility when a manuscript receives both a positive and a negative review. Since some of the negative reviews were incorrect or inappropriate in terms of the manuscripts data, editors may need to provide some mechanism to address this issue.

Miller and Serezan (1984) believe that editors can provide more direction regarding the submission of manuscripts. They maintain that "Although the responsibility for selecting a journal for publication is clearly that of the author, the selection is more precise if the journal publishes guidelines which define its purpose, scope and audience" (p. 678). Bowen, Perloff, and Jacoby (1972) also emphasize the role of the editor and state that "Publications of the standards and procedures to be used (in the review process) would help authors submit better articles in the first place" (p. 224).

Finally, the findings provide insight into the people behind the comments — the reviewers. While it is not suggested that each reviewer easily fits a nice profile or slot, reviewers might introspect about their orientation to reviewing manuscripts. Diverse reviews are more likely to occur when, for example, a manuscript is reviewed by a competitor and a critic who recommends acceptance, a conciliator and a critic who recommends rejection, or especially by a conciliator and a competitor or procrastinator. A review outcome may depend on the luck of the draw as in the case of having a manuscript reviewed by a competitor and a procrastinator. The reviewers have a responsibility to examine their own stance for, as Scott (1974) points out, even the use of specific guidelines will not overcome reviewer bias. The categories devised for this paper may provide insights to editors for that purpose.

Mahoney (1977) argues that peer review is a form of evaluative testing. He further states that if a psychological instrument were at focus, educators/researchers would demand that its validity and reliability be demonstrated. Indeed, reviewers quite often reject manuscripts on the basis of the lack of such data. Yet, ironically, reviewers are part of a testing procedure for which validity and reliability have not been demonstrated. Mahoney asks, "Have we assumed that scientists are somehow unaffected by the processes which appear to be so common in other members of the process?" (p. 173). Nevertheless, many would agree with Wilson (1978) that while peer review needs improvement, it serves a worthwhile function in the publication process. Banner (1988) agrees with Wilson but is more specific in his conditions for its merits. He believes that peer review is valuable, "but only if its structure and operations are repeatedly tested and reviewed and only if its custodians are scrupulous and demanding" (p. 112).

References

Banner, J. (1988). Preserving the integrity of peer review. *Scholarly Publishing, 19*, 109-114.

Bowen, D., Perloff, R., & Jacoby, J. (1972). Improving manuscript evaluation procedures. *American Psychologist, 21*, 221-225.

Commoner, B. (1978). Peering at peer review. *Hospital Practice, 13*, 25, 29.

Ingelfinger, F.J. (1974). Peer review in biomedical publication. *The American Journal of Medicine, 56*, 686-692.

Mahoney, M.J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*, 161-175.

Marsh, H., & Bell, S. (1981). Interjudgmental reliability of reviews for the Journal of Educational Psychology. *Journal of Educational Psychology, 73* 872-880.

Miller, A.C., & Serezan, S.L. (1984). Criteria for identifying a refereed journal. *Journal of Higher Education, 55*, 673-699.

Scott, W.A. (1974). Inter-referee agreement on some characteristics of manuscripts submitted to the Journal of Personality and Social Psychology. *American Psychologist, 29*, 698-702.

Wilson, J.D. (1978). Peer review and publication. *The American Society for Clinical Investigation Inc, 61*, 1697-1701.

Zukerman, H., & Merton, R. (1971). Patterns of evaluation in science: Institutionalism, structure, and functions of the refereeing system. *Minervax, 9*, 66-100.