Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation

J. Scott Armstrong, *The Wharton School, University of Pennsylvania, USA*

Keywords: blind reviews, early acceptance, electronic publishing, innovation, post-publication peer review, results-blind reviews

ABSTRACT: *This paper reviews the published empirical evidence concerning journal peer review consisting of 68 papers, all but three published since 1975. Peer review improves quality, but its use to screen papers has met with limited success. Current procedures to assure quality and fairness seem to discourage scientific advancement, especially important innovations, because findings that conflict with current beliefs are often judged to have defects. Editors can use procedures to encourage the publication of papers with innovative findings such as invited papers, early-acceptance procedures, author nominations of reviewers, structured rating sheets, open peer review, results-blind review, and, in particular, electronic publication. Some journals are currently using these procedures. The basic principle behind the proposals is to change the decision from whether to publish a paper to how to publish it.*

Formal peer review is used to make decisions about who should receive grant money, who should be hired or promoted, and which papers should be published. A common thread here is the use of peer review to allocate scarce resources. This paper examines the use of formal peer review to decide how to allocate limited space in journals.*

This author's impression is that most successful researchers find the current system of journal peer review to be effective. Journal peer review is commonly believed to reduce the number of errors in published work, to serve readers as a signal of quality, and to provide a fair way to allocate journal space. As an editor, referee, and author, I cannot recall a single instance in which a journal accepted a paper upon its first submission, without making suggestions for change. The changes almost always have led to improvements, although sometimes the level of effort and delays seemed excessive when judged against the benefits.

* Horrobin (1996) reaches similar conclusions in his discussion of peer review for grant proposals.

Address for correspondence: J. Scott Armstrong, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. The author, a professor of marketing at the Wharton School since 1968, was a founder editor of the *Journal of Forecasting* and the *International Journal of Forecasting*.
This paper is based on a presentation at a workshop, "Advances in Peer Review Research", American Association for the Advancement of Science Meeting, Baltimore, MD, February 9, 1996.
Paper received, 15 April 1996; revised, 4 November 1996; accepted, 15 November 1996.
ISSN 1353-3452

The primary concern of this paper is with the publication of innovative findings, or more precisely, the lack thereof. By innovation, I mean useful and important new findings that advance scientific knowledge. Such findings typically conflict with prior beliefs, for which Kuhn (1962) uses the term "paradigm shift". This concern on innovation is not often discussed directly in the literature about peer review. In Stamps' (1997) classification of publications on peer review, only his 12th ranked topic, "conservatism," addressed issues related to innovative findings.

Firstly, I describe the procedures used to search for and interpret empirical findings on peer reviews. In this examination, I consider quality control, fairness to authors, and innovative findings. I then provide suggestions, mainly for editors, but also for reviewers and authors.

PROCEDURES FOR THE LITERATURE REVIEW

I tried to review all published empirical literature on peer review. Searching for such studies is difficult because they are published in a variety of disciplines such as medicine, sociology, psychology, physics, engineering, social science, management science, and economics. Furthermore, the terminology that authors use in peer review varies, so it is difficult to locate relevant papers. The term "peer review" often does not appear in the titles, and the titles for many references in this paper would fail to alert the searcher.

Initially, I used references from key papers and books, such as Kupfersmid & Wonderly (1994). Then I circulated drafts of this paper to researchers; they told me about 43 additional relevant papers.

Continuing the search and following Stamps' (1997) sources and materials, I examined Speck's (1993) annotated review. While he did not describe the procedure used to select studies, his search, which covered 1960 through 1991, was extensive. He found 643 academic papers on journal peer review. I coded his summaries and concluded that 101 of the papers provided empirical evidence. Some of these empirical studies were excluded because they were not relevant to the issues in this paper. Speck's list and my review overlapped somewhat. Of the 38 empirical papers published by 1991 that I found, Speck cited only 23, but he listed 13 that I had overlooked. I located 17 papers published since Speck's book, bringing the total to 68. These empirical studies are denoted by "E" at the end of the references. Some of these papers are reviews of several empirical studies, so the total number of empirical studies exceeds the number of papers cited here.

The empirical peer review literature is of recent origin. All but three of the 70 studies were published since 1975. Papers reporting on experimental (or quasi-experimental) studies are of particular interest: twelve exist and they tend to be controversial. These studies are denoted by "X" at the end of the references.

Most of the empirical research focused on fairness and quality; researchers have paid little attention to innovation. For example, in coding the 101 empirical studies annotated by Speck, I found only four directly related to innovation.

In March 1996, to help ensure that my interpretations of the studies were correct, I sent a draft of this paper to all those who had published empirical research on this topic, providing that recent addresses were available. I contacted the lead author if an address was available, otherwise a co-author. I did not contact authors of papers that were discovered and added to the list after this survey. In the letter I asked "Do I summarize your work properly, and if not, what changes should I make?" and "Are you aware of any empirical work that I should include?" I heard from authors of 32 papers (counting five by myself and co-authors). I mark these papers with "R" at the end of the reference. Typically the authors agreed with my interpretations, but occasionally they made suggestions to clear up minor errors of interpretation.

QUALITY CONTROL

Here is how the current quality control system works. Researchers, sometimes working in teams, spend hundreds of hours working on a specialized topic, often collecting empirical evidence and applying formal analytical techniques. They write papers and often benefit from pre-submission peer reviews. They strive to follow standards for scientific work and they sign their names to their work. Their futures depend to some extent on the quality of their paper.

These papers are then reviewed by people who are working in related areas but generally not on that same problem. So the reviewers typically have less experience with the problem than do the authors. Of course, there may be aspects of the research, such as methodology, in which the reviewers have more expertise.

Reviewers generally work without extrinsic rewards. Their names are not revealed, so their reputations do not depend on their doing high quality reviews. Perhaps this leads them to spend little time on their reviews. In any event, on average, reviewers spend between two and six hours in reviewing a paper (Jauch & Wall, 1989; King, McDonald & Roderer, 1981; Lock & Smith, 1990; Yankauer, 1990), although they often wait for months before doing their reviews. They seldom use structured procedures. Rarely do they contribute new data or conduct analyses. Typically, they are not held accountable for following proper scientific procedures. They match their opinions against the scientific work by the authors.

Reviewers' recommendations often differ from one another, as shown by Cicchetti (1991). Most authors have probably experienced this. For example, here are reviews for one of my papers: Referee #1: " . . . The paper is not deemed scientific enough to merit publication." Referee #2: " . . This follows in the best tradition of science that encourages debate through replication."

Authors are critical of the quality of the reviews that they receive. Bradley (1981) asked authors about their experience on the "last compulsorily revised article published in a refereed journal". When asked whether the changes advocated by the referees were based on "whim, bias, or personal preference", only 23% said none were, while 31% said that this applied to some important changes. Forty percent of the respondents said that some of the referees had not read the paper carefully.

Given this background, it is interesting that editors decide whether to publish a paper primarily based on two or three reviews. Authors may appeal the decision, and some journals have formal procedures for appeal, but I suspect that editors rarely change decisions.* The editors of the *American Sociological Review* agreed with the

* I consider myself fortunate in this regard as editors have over-ruled referees for most of my important papers. When I went through my list of refereed journal articles and picked what I thought to be the 20 most important papers, I found that none of them had received all favorable reviews. (In many of these cases I urged the editors to publish the negative reviews and my response.) On the other hand, about half of the 20 next most important papers had all favorable reviews.

authors on only 13% of the decisions that were appealed (Simon, Bakanic & McPhail, 1986).

Using peer reviewers to control quality seems most appropriate for papers that contain little that is new. If researchers are pioneering new areas, or if their findings depart from what reviewers currently know in an area, reviewers will have difficulties. To the extent that findings are new, the expertise of the reviewers is likely to be less than that of the authors.

A distressing aspect of the current quality control system is that work by the best researchers is, on average, judged by those who are less capable. This occurs if editors pick "fairly" (at random) from among potential reviewers. Of course, some editors may ask their best reviewers to examine papers by authors with good reputations.

False Cues

Reviewers appear to base their judgments on cues that have only a weak relation to quality. Such cues include (1) statistical significance, (2) large sample sizes, (3) complex procedures, and (4) obscure writing. Researchers might use these cues to gain acceptance of marginal papers (Armstrong 1982).

Although it typically has little relationship to whether the findings are important, correct, or useful, *statistical significance* plays a strong role in publication decisions as shown by studies in management, psychology, and medicine (Begg & Berlin, 1988; Greenwald, 1975; Hubbard & Armstrong, 1992; Salsburg, 1985; Sterling, Rosenbaum, & Weinkam, 1995). The case against statistical significance is summarized for psychologists by Cohen (1994) and for economists by McCloskey & Ziliak (1996). If the purpose is to give readers an idea of the uncertainty associated with a finding, confidence intervals would be more appropriate than significance tests.

Atkinson, Furlong & Wampold (1982) conducted an experiment to determine whether reviewers place too much emphasis on statistical significance. They prepared three versions of a bogus manuscript where identical findings differed by the level of statistical significance. The reviewers recommended rejection of the paper with nonsignificant findings three times as often as the ones with significant findings. Interestingly, they based their decision to reject on the design of the study, but the design was the same for all versions.

Using significance tests in publication decisions will lead to a bias in what is published. As Sterling (1959) noted, when studies with nonsignificant results are not published, researchers may continue to study that issue until, by chance, a significant result occurs. This problem still exists (Sterling, Rosenbaum & Weinkam, 1995).

Large sample sizes are used inappropriately. Sometimes they are unnecessary. For example, reviewers often confuse expert opinion studies with surveys of attitudes and intentions. While attitudes and intentions surveys might require a sample of more than a thousand individuals, expert opinion studies, which ask how others would respond, require only 5 to 20 experts (Armstrong, 1985, p. 96). Even when sample size is relevant, it is likely to be given too much weight. For example, Lau (1994), in a study of election polls for the U.S. presidency, concluded that the sample size of the surveys was loosely related to their accuracy.

Complex procedures serve as a favorable cue for reviewers. One wonders whether simpler procedures would suffice. For example, in the field of forecasting, where it is possible to assess the effectiveness of alternate methods, complex procedures seldom help and they sometimes harm accuracy (Armstrong 1985). Nevertheless, papers with complex procedures dominate the forecasting literature.

Obscure writing impresses academics. I asked professors to evaluate selections from conclusions from four published papers (Armstrong, 1980). For each paper, they were randomly assigned either a complex version (using big words and long sentences, but holding content constant), the original text, or a simpler version. The professors gave higher ratings to authors of the most obscure passages. Apparently, such writing, being difficult to understand, leads the reader to conclude that the writer must be very intelligent. Obscure writing also makes it difficult for reviewers and readers to find errors and to assess importance. To advance their careers, then, researchers who do not have something important to say can obfuscate.

Effects of Journal Peer Review on Quality

Journal peer review improves the quality of a given paper. In the author survey by MacNealy, Speck & Clements (1994), 80% of the 96 authors responding said that they found the reviewers' suggested revisions to be "reasonable". Bradley (1981), in a survey of 361 statisticians and psychologists, found that 72% thought that "The net effect of refereeing upon the quality of the article was to improve it." Fletcher & Fletcher (1997) report on an experiment comparing medical research manuscripts as they were received with the same manuscripts after being revised based on reviewing and editing. Raters (who were blind to the treatment) concluded that the revised papers were superior on 33 of 34 elements of quality. Surprisingly, Fletcher & Fletcher report on another study concluding that editing did little to improve readability.

Although peer reviewing improves the quality of papers, its value in making comparisons across papers is less obvious. Gottfredson (1978), using from one to three experts per paper, found that their ratings of the "quality" of 387 published psychology papers had a weak relationship to the number of times they were cited over an eight-year period.

Although journal peer review improves quality, it does not assure it. It seems likely that errors can be found in any published paper. Consider such a simple measure of quality as the accuracy of references. Evans, Nadjari & Burchell (1990) studied the references cited in three medical journals and found a 48% error rate. They said that "a detailed analysis of quotation errors raises doubt in many cases that the original reference was read by the authors." Eichorn & Yankauer (1987) found that 31% of the references in public health journals contained errors, and 3% of these were such that they could not locate the references. In addition and more important, Eichorn and Yankauer found that the authors' descriptions of previous studies differed from the original authors' interpretations for 30% of the citations, with half of these descriptions being unrelated to the authors' contentions.

Horrobin's (1982) opinion, as editor of two biomedical journals, was that only one-third of reviews are accurate. In an analysis of 216 reviews for 108 manuscripts on reading and language, Fagan (1990) concluded that reviewers made mistakes.

Murray (1988) evaluated the statistical procedures described in 28 papers that had been published in the *British Journal of Surgery* and concluded that four papers should have been rejected, seven needed major revisions, and 11 needed minor changes. Thus, only 21% of the papers passed this post-publication peer review without the need for changes.

Stewart & Feder (1987) provide further evidence of errors in published papers. They studied the 18 papers published by John Darsee between 1978 and 1981, shortly before he admitted to scientific fraud. The papers had been reviewed and published in major biomedical journals. Stewart and Feder found errors in 16 of the 18 papers. On average, they found 12 errors per paper, some minor, but many were major. For example, it was reported that the father in one family had his first child at age eight and the next at age nine.

The reviewers' ability to identify even such major transgressions as plagiarism seems weak. Epstein (1990) conducted an experiment whereby he sent out two modifications of a previously published paper that were intentionally flawed methodologically. They were sent to journals in social work, sociology, psychology, counseling, and medicine. Only two of the 110 journals to which he sent it said that the paper had already been published. This occurred despite the fact that the paper had been cited frequently. Although the study's control group had been omitted from the original paper, few reviewers mentioned this as a problem. Epstein concluded that only six of the 33 reviews that were received were competently done.

One can find evidence of errors by looking at replications. If an original study and a replication reach different conclusions, then one is likely to be in error. How often does this occur? Hubbard & Vetter (1996), in their study of 266 replications and extensions in accounting, economics, finance, management, and marketing, found that 46% of them conflicted in a major way with the original findings, and an additional 27% conflicted to a minor extent. While this suggests a problem, we cannot infer the extent of the problem because editors might be biased against or in favor of publishing failed replications, or authors might tend to replicate studies that are excellent or that are flawed.

In a study of papers published in the *Journal of Money, Credit and Banking*, Dewald, Thursby & Anderson (1986) obtained computer programs and data from authors of studies published in economics and attempted to replicate the results. They found that errors in transcription, data transformation, and computer programming were commonplace and that some of these errors significantly changed the statistical results and conclusions of the studies examined.

FAIRNESS

Given that publication is so important in academic hiring and promotion decisions, the issue of fairness receives much attention. Almost half of the empirical papers on journal reviewing listed by Speck (1993) address it.

One belief is that authors should receive the same treatment as one another. This does not always happen. Peters & Ceci (1982) conducted a quasi-experiment in which they concluded that reviewers were biased against authors from unknown or less-prestigious institutions. The results of this experiment were surprising. When I showed this design to a group of 22 academic researchers as a proposed study, they did poorly at predicting the results (Armstrong, 1982). For example, they expected that 46% of the studies would be rejected as "adding nothing new". In fact, none were.

Lloyd (1990) conducted an experiment in which she sent for reviews on the identical manuscript on psychology with names that were obviously male or female (all fictitious). She found that female-authored manuscripts were accepted significantly more often by female reviewers (62%) than by male reviewers (21%).

Journals clarify and standardize the rules to try to ensure that all authors get the same treatment. One apparent aspect of fairness is that the opinions of reviewers, rather than the editor, should determine the disposition of the paper. But with this policy, the decision is heavily influenced by people who have little concern for the ability of the journal to interest its readers.

Editors who are concerned about fairness often treat referees' recommendations as votes. In a fair procedure, then, a paper with all positive reviews should be published and one with all negative reviews should be rejected. The papers with mixed reviews can then be ranked on the proportion and strength of favorable reviews. For the most prestigious journals, where the supply of papers is high, the journal is likely to be filled with papers that have all positive reviews. Kupfersmid & Wonderly's (1994, p. 56) summary of four empirical studies led them to conclude that papers that received mixed reviews have a low probability of being published. Three additional studies are relevant: In a study of 263 blind reviews for the *Journal of Counseling Psychology* from 1982 through 1983, Munley, Sharkin & Gelso (1988) found that the editors' decisions to publish were highly correlated with reviewers' recommendations. Marsh & Ball (1989) obtained similar results. Bakanic, McPhail & Simon (1990) evaluated the manuscripts submitted to the *American Sociological Review* from 1977 to 1982 and found that "a single recommendation to reject often resulted in a rejection".

One approach to improving fairness has been to use more reviewers (Cicchetti, 1991). This would make it more difficult to publish innovative papers in the leading journals because it is more likely that a reviewer would provide a negative review. Interestingly, the lack of reliability in reviews should improve the chances for the publication of papers with innovative findings.

Blind Reviewing

Double-blind reviews, refered to here as 'blind reviews', are used to help ensure that papers by unknown researchers or by researchers from less prestigious institutions are reviewed fairly. Blind reviewing is popular, as shown by Rowney & Zenisek's (1980) survey of researchers in psychology.

The evidence on the effect of blind reviewing is mixed. Blank (1991) conducted a randomized experiment on peer review for economics journals and found that the acceptance rates of submissions by authors at top-ranked and at low-ranked universities did not differ much whether or not the reviewing was blind.

Fletcher & Fletcher (1997) report on experiments with blind review. In one study of 123 papers on medicine, blinding was successful in shielding the authors' identity for 73% of the papers. Reviews by blinded reviewers were judged by the editors to be higher in overall quality. The authors of the papers judged the blinded reviewers to be more knowledgeable than those where the authors' identities were known. They also thought the blinded reviewers were fairer.

After studying peer reviews for the *Journal of General Internal Medicine*, McNutt, Evans, Fletcher & Fletcher (1990) found little difference between blinded and nonblinded reviews with respect to recommendations about publication. However, the editors who rated the reviews thought that the nonblinded reviews were more constructive and more courteous, and authors rated them as fairer.

Blind reviewing discards important information: namely, whether the researcher has been successful in prior efforts. This is useful for predicting the impact of a paper. For example, Abrams (1991), who looked at research by ecologists, found that citations for a researcher's publications were highly correlated to the citation rates achieved by his or her previous papers.

If reviewers were biased in favor of those at prestigious institutions, then these papers should be of lower quality than those from less prestigious institutions. We have no evidence that this occurs. In fact, Perlman (1982) studied 120 papers from two psychology journals and found that papers by authors at institutions with higher status were cited much more frequently than papers by authors at lower-status institutions. Thus, Perlman concluded that blinding harms quality.

An insistence that referees should not know the authors of papers that they review has an unfortunate consequence. It discourages researchers from seeking informal peer review from those who are probably best qualified to help them, because the researchers may not want them to be ruled out as referees.

Some management journals practice an extreme form of blind review. They request that referees disqualify themselves if they think they know who the authors are. This poses a problem for well-respected authors because competent researchers in the field will recognize their work. Only those ignorant of the literature would be able to provide reviews for leading researchers.

INNOVATION

By innovative studies, I mean those with evidence that existing beliefs are incorrect. More generally, innovation refers to all advances in scientific knowledge. I am concerned with important innovations; trivial new findings threaten no one and are unlikely to offend reviewers.

At first glance, one might wonder whether innovation poses a problem. I think that we are all in favor of publishing innovative papers. Indeed, a survey of journal reviewers in 43 disciplines found that "originality" was the first-ranked consideration in reviewing papers (Juhasz, Calvert, Jackson, Kronick & Shipton, 1975). In a survey of Canadian psychologists, Rowney & Zenisek (1980) found that they would evaluate favorably papers with a "new, original theory". Kerr, Tolliver & Petree (1972) obtained similar findings in their survey of board members for 19 journals in management and social sciences. Beyer, Chanove & Fox (1995) found that 18% of the authors they surveyed claimed novelty in their papers, and this was correlated positively (though weakly) with the final decision on acceptance. It is unlikely, then, that there is an *intentional* bias against innovation.

Kuhn (1962) claimed that when innovative findings conflict with important beliefs, resistance is likely to be strong and long lasting. Barber (1961) describes the fierce resistance met by famous scientists. Resistance seems endless for some important issues. For example, protestors said that Rushton should not be allowed to present his findings on brain size and intelligence (Rushton & Ankney, 1996) at the AAAS meetings in Baltimore in February, 1996. The protests were so strong it was deemed necessary to provide security guards. This occurred even though similar "undesirable" findings have been published since the late 1800s,

Horrobin (1990) examined important new findings in medical science and concluded that journals' emphasis on quality control often led to extensive delays in their publication. He also concluded that the stress on quality leads reviewers to ignore the importance of the study.

Reviewers tend to reject papers that provide important new findings. Experimental studies in psychology support this viewpoint. For example, Goodstein & Brazis (1970) asked 282 psychologists to review one of two abstracts that were identical except for the results. The psychologists rated those in which the results were in accord with their own beliefs as better designed and said that they were more suitable for publication. Mahoney (1977) showed that reviewers rejected papers with controversial findings because, they said, the methodology was poor, while they accepted papers with identical methodology that supported conventional beliefs. Experiments by Abramowitz, Gomes & Abramovitz (1975) and Epstein (1990) yielded evidence that was consistent with Mahoney's. In laboratory experiments with graduate students and practicing scientists, Koehler (1993) found that, when given results on a controversial issue, they rated the quality of a research report higher if it agreed with their prior beliefs. In a nonexperimental study, Smart (1964) found that 30% of studies in doctoral dissertations in psychology failed to support a dominant hypothesis, but this dropped to 20% of papers presented at the annual American Psychological Conference, and to 10% of papers published in journals.

In a survey of the editors of 16 leading American Psychological Association journals, Armstrong & Hubbard (1991) found that papers with controversial findings were harshly reviewed. For the two-year period covered by the survey, only one paper with controversial findings received favorable reviews from all reviewers. The editor revealed that he had been determined to publish the paper, so he had sought referees that he thought would provide favorable reviews. This represents one favorably reviewed paper with controversial findings for 32 'journal years'.

Anecdotal evidence also suggests that it is common for reviewers to reject ground-breaking papers. Gans & Shepherd (1994), in their survey of famous economists, conclude that the leading journals stifle innovative work. Garcia (1981) provides similar examples in psychology. Campanario (1995) describes examples in which "citation classics" were rejected by the journal to which they were first submitted. He concludes that "something must be wrong with the peer review system when an expert considers that a manuscript is not of enough interest and it later becomes a classic in its discipline" (p. 321).

While one could argue that innovations should be subject to more careful scrutiny and that they will eventually be published anyway, it is just as easy to develop arguments why such studies should be published more quickly. Given the current situation, the barriers might be so great that researchers become discouraged and decide it is more rewarding to focus on their own advancement rather than the advancement of science. Why invest time working on an important problem if it might lead to controversial results that are difficult to publish?

POSSIBLE CHANGES IN JOURNAL REVIEWING

The following changes in peer review might encourage the publication of innovative work. They are intended to lead to improvements in papers and to avoid the rejection of papers with innovative findings. The suggestions are arranged roughly in the order of the decisions in the reviewing process.

Invited Papers

Given that peer reviewers pose barriers for innovations, journals might reduce peer reviewing as a screening device. One way is to make greater use of invited papers. This could be done for a section of a journal or for at least one journal in a field. For example, the *Journal of Economic Perspectives* (*JEP*) asks well-established researchers to address important problems. These invited papers usually receive informal peer review (to improve quality) at the request of the author or the editor. While the *JEP* publishes some papers that it does not solicit, this is unusual (many are submitted, but few are accepted). The *JEP* has become highly regarded and its papers are often cited.

Rodman & Mancini (1977) surveyed editors of journals in education and found the practice of inviting papers to be widespread; 89% of the 28 editors who responded had published "inside track submissions".* After surveying papers published in six leading technical communication journals, MacNealy, Speck & Clements (1994) concluded that editors had solicited over one-fourth of their published papers. When inviting papers, editors can rely heavily on researchers who have a good track record and who are presumably interested in maintaining their reputations. Editors can also encourage and assist authors in obtaining peer review.

Unfortunately, I could find no evidence on whether invited papers are better or worse than those that go through normal channels. One might, for example, compare citation rates for papers from each category.

Early Acceptance Procedures

Editors could accept papers based only on a review of the research design. This might encourage submission of papers that will produce innovative findings. Given acceptance, the researcher would then have to carry out the design as specified. The eventual paper would be published irrespective of the results. Presumably, then, editors would want to accept studies that would be interesting no matter how the results turned out.

With an early acceptance procedure, researchers could find out whether it was worthwhile to do research on a controversial topic before they invested much time. An additional benefit of such a review is that they receive suggestions from reviewers before doing the work and can then improve the design. A paper accepted under this procedure, when completed, would again be sent for review to improve quality. Revisions might be requested, say for the writing, but the paper would still be considered as accepted.

* Rodman & Mancini (1977) concluded that the practice of inside submissions is unethical and it should be banned. They did not examine the impact that this might have on the communication of scientific findings.

Weiss (1989) used an "advance publication review" procedure for *Applied Psychological Measurement*. He expected that this procedure would improve efficiency by allowing researchers to work on projects that would be published. He ended this trial in 1996 for lack of interest (personal communication). Authors had submitted only five papers under this procedure and they did not tackle topics that were controversial; Weiss speculates that this is because the journal specializes in methodology and controversy is rare in this field.

Simultaneous Submissions

Currently, many journals review papers on the condition that they are not being reviewed elsewhere. Simultaneous submissions would speed up the reviewing process, because the journal that is first to accept is likely to gain the paper for publication. This would reward efficiency in the operation of journals. For a favorable view of this proposal, see Szenberg (1994); for an unfavorable one, see Pressman (1994); and, for a balanced view, see Lindsey (1978).

Authors often submit scientific books to several publishers simultaneously. A trial of this procedure might help journals to assess such questions as: Are there enough reviewers? Would journals act in haste? Would authors take the easy road and avoid revisions that could greatly enhance the value of their research? One way that this could begin would be for two or more journals to allow for simultaneous submissions. If this works well, other journals might join in the agreement.

Pre-submission Reviews

In their instructions to authors, journals might state a preference for papers that have been reviewed by the authors' peers before submission. Reviews by established researchers, presentations at research conferences, and publication in proceedings would be favorable signs for those concerned about quality.

Nomination of Reviewers by Authors

Researchers usually know who has expertise on their topic. Therefore, journals might encourage them to nominate reviewers for their paper, particularly for papers with controversial findings. Some journals, including *Science, Organization Science, Personality and Social Psychology Bulletin* (Hendrick, 1976), and the *Journal of Molecular Biology* invite authors to suggest reviewers.

Results-blind Reviews

Given the evidence of bias against papers with controversial findings, researchers in some disciplines have suggested results-blind reviewing (*e.g.*, Newcombe, 1987, recommends it for medical research). Here, authors are invited to provide a version of the paper with the results and conclusions omitted. If the editor believes that the paper is concerned with an important topic, he would give the reviewers a version with results and conclusions omitted. This procedure should also help to reduce reviewers' bias against papers that do not have statistically significant results and to reject statistically significant but unimportant studies.

The *International Journal of Forecasting* currently allows for results-blind reviews. One would expect that such a procedure would rarely be needed. Not surprisingly, few authors have used this option. The intent, however, is to have it available for the few cases where it might help. Also, the availability of this option helps to protect editors against claims that they are arbitrarily biased against new findings.

Structured Reviewer Forms

Because reviewers often use false cues, journals should explicitly describe the cues that should be evaluated. Structured rating sheets can emphasize a preference for papers having comparative empirical studies, important topics, and surprising results. The structured rating sheet might also tell reviewers what to ignore. The American Psychological Association is considering the elimination of significance tests from papers submitted to its journals because they often serve as a false cue of quality (Shea, 1996).

Reviewers appear to seek reasons why a paper should not be published. The presumption behind this quality control system is that it is important to protect other scientists from encountering flawed work. To knowingly publish a paper with major errors is regarded as scientific misconduct. However, Chalmers (1990) suggests that the failure to report important findings might also be regarded as scientific misconduct. This seems obvious, for example, if a researcher obtained findings that a currently used medical treatment has serious undesirable side effects that had not previously been realized. This principle also applies to other areas. For example, what if a widely used procedure in management is harmful? I obtained such evidence in a study of the Boston Consulting Group (BCG) product portfolio matrix. This was the first experimental evaluation of a technique that had been widely used for almost a quarter century. My paper went through seven rounds of reviews, with fourteen referees, at four journals before it was accepted after a three-year reviewing process. While the reviewing led to additional experiments and to improvements in the writing, no errors were discovered and no substantive changes occurred in the conclusions (Armstrong, 1996). To counter a bias of looking more closely for errors when there are important new findings, it might be worthwhile to structure the reviewer form so that it asks why a paper *should be* published. The *International Journal of Forecasting* asks this on its reviewer's form.

Given an apparent bias toward papers that support existing beliefs, authors might be encouraged to report evidence that opposes their findings. Referees can examine whether the author provides a complete and unbiased review of prior research. For example, I (Armstrong, 1996) found that the ten researchers who reported confirming results in their extensions of escalation bias (which states that managers invest more heavily if a project goes bad) failed to cite any of the four published papers that contained contrary results.

To avoid acceptance decisions based on voting and to encourage ideas about improvements, journals should not ask reviewers for recommendations about whether the paper should be published. Instead, the rating form would direct reviewers to suggest ways to enhance quality. The editor-in-chief would make the publication decision, or in broad fields, associate editors could be given this responsibility. The editors must decide which papers are relatively most important.

Open Peer Review

One way to seek innovation is for editors to publish controversial papers along with comments by reviewers, as is done by *Behavioral and Brain Sciences* (Harnad, 1979). A section editor could be allocated space for finding papers with innovative findings and publishing them along with commentary. The risk of publishing a paper that could be in error is balanced by providing the readers with commentary.

Journals can encourage open review by providing space for replication studies, corrections of errors, commentaries on previous papers, and letters to the editor. These procedures are used by some prestigious journals such as *Science*, *Nature*, *Journal of Economic Perspectives*, and the *American Psychologists*. In some fields, such as management, journals make little use of these procedures. Furthermore, Hubbard & Armstrong (1994) found that the proportion of published comments in leading marketing journals declined by one-third between 1974-79 and 1980-89.

Journal policies can make a difference. The *Quarterly Journal of Business and Economics* announced in 1984 that it would give priority to the publication of replications and extensions and this led to a substantial increase in the number of replication studies that they publish (Fuess, 1996).

Journal policies should give special attention to procedures for informing the scientific community about major errors, mistakes, or fraud in journal articles. Pfeifer & Snodgrass (1990) studied 82 papers that had been published between 1973 and 1987 but were later retracted. They concluded that these papers were not being purged from the literature effectively, as they continued to be cited favorably. Friedman (1990), who studied fraudulent research by Slutsky, concluded that existing procedures made it difficult to track corrections and that some journals fail to publish corrections. Major errors can have a continued existence long after publication, as shown by such well-known examples as the Cyril Burt case (Wade, 1976) and others (Broad & Wade, 1982).

Alternative Formats

Editors often justify their use of peer review for screening as the best way to allocate scarce space in journals. In many fields, however, journals devote space not to reporting on scientific findings but to researchers' "proposals". In management and social science journals, for example, a large percentage of the papers contain elaborate descriptions of models or methods or phenomena. These are presented with logical arguments or with complex mathematical descriptions. The authors conclude with suggestions that someone should test their proposals. Seldom does another researcher take up the challenge. Journals could free up a lot of space if they announced that they would publish proposals only after they have been tested. The papers could then focus on presenting findings. They could put important technical details in small print in an appendix, or they could be put on file with the editors and interested readers could write for details. For example, *Nature*, as a condition of publication, requires that the materials and methods used in the study be "freely available to academic researchers for their own use". (Perhaps you can estimate how many researchers would request details for the typical paper published in your field.) Should the demand for such details be large, the details could be published. One promising option is to make details available through electronic publication, as is being done by some journals, such as *Interfaces*.

If the focus is on scientific advancement and communication, editors might change their orientation from deciding *whether* to publish a paper, to determining *how much* of the paper should be published. This could range from publishing only the title, to an abstract, an extended abstract, a note, a short paper, or a long paper. In each case, information would be provided about how to get more details, perhaps using the Internet. *Nature* allows for five types of publications ranging from peer-reviewed “scientific correspondence” (500 word maximum) to “review articles” (six page maximum). *Science* offers eight different categories.

Electronic publishing by journals will do much to solve the problem of scarce resources. It will become inexpensive to disseminate papers. Electronic publishing may gradually supplant paper journals as a primary channel for innovative findings. It can reduce time lags and make the findings more accessible. Formal editorial boards can review these papers and, given the low cost of on-line space, they could recommend publication along with a summary judgment to guide readers and evaluators. Authors could then decide whether to publish the paper, summary judgment, and reviews, or to also include a reply to the reviewers, or to withdraw the paper.

Electronic publication should also aid in post-publication review. It would be inexpensive to include peer reviews. Readers can publish comments on papers. Records might be kept of the number of times a paper is accessed and how many errors are found. For an example of such a publication on the Internet, see Hibbitts (1996a), which addresses issues of peer review in the legal profession. Hibbitts’ electronic paper offers advantages over its paper version.*

Hibbitts (1996b) discusses various aspects of electronic publication for law reviews.** He discusses the resistance to electronic publication and draws an analogy to the negative reaction of many scholars when the printing press was introduced. Despite some resistance, journals in physics (Taubes, 1994), biology, management, and other areas have begun to use electronic publishing in various ways.

ACTIONS BY EDITORS, REVIEWERS AND AUTHORS

The selection of an editor is perhaps the most important decision that a journal makes. Franke, Edlund & Oster’s (1990) analysis of 17 management journals over a 12-year period led them to conclude that a journal was more successful if the editor was a successful researcher. Based on the discussion above, journals should avoid the selection of bureaucrats.

In order for innovative papers to be published, editors should take an active role. Roediger (1987) suggests that editors can have a strong influence by encouraging the

* Although we do not know the numbers of readers of the paper version of Hibbitts’ (1996a) paper, it is likely to be only a fraction of those who read the electronic version. In its first year, the electronic version was accessed on over 3,200 occasions.

** To be sure, law reviews differ in many ways from scientific journals. Perhaps the most obvious difference is that the editors are students.

submission of papers that would otherwise be unseen or unwritten.* He refers to the editor as a scientific scout who would seek papers with innovative findings and find reviewers who will provide constructive and objective reviews. Some journals actively seek to publish controversial papers. These include *Behavioral and Brain Sciences*, *IEEE* (Christiansen, 1978), the *International Journal of Forecasting*, and *Psychological Bulletin*.

Editors may need to take risks on important issues. Arkes (1996) suggests that it may help journals to publish innovative papers if a board of editors makes decisions rather than a single editor. In addition, he suggests that reviewers be told in advance that their reviews would be examined by the board and be informed about who wrote the reviews. A study by Tetlock & Kim (1987) suggests that such accountability is likely to improve the quality of judgments.

Editors need to judge the reviewers' work and to hold reviewers to the same standards of science that they expect of authors. Liversidge (1989) presents an example, where he, as editor, published a controversial paper about AIDS, because although the reviewers were negative, they could not support their position. In addition, authors should be permitted to review the reviewers. If errors are noted in the reviews, consideration should be given for additional reviews. Such a policy is informally adopted by many journals, and *Science* has this as an explicit policy.

Even if editors do not rely on reviewers for recommendations, they can still use them to help ensure that papers are fairly free of errors. They should guard against biased research findings by requiring that authors and reviewers report funding sources and potential sources of bias (Chalmers, Frank & Reitman, 1990). Davidson (1986), Cho & Bero (1996), and Needleman (1992) show how such bias affects findings in medical research.

Much of the discouragement of innovative work stems from reviewers. Reviewers can avoid this by adopting a role to try to improve papers. In doing so, they would not make a recommendation as to whether the paper should be published. In my own reviewing of papers, I no longer make recommendations about acceptance. I thus remove myself as a barrier to publication and do not reject my colleagues' papers. The decision rests on the editor, who can choose from among many papers without being constrained by my "vote". This practice fits with my policy as a reviewer of generally revealing my identity to authors.

Some journals solicit innovative work, but few researchers avail themselves of this opportunity. If authors are determined to work on topics that might challenge existing beliefs, they now have the opportunity to self-publish electronically. Without certification from a journal, however, it may be difficult to attract a readership. Readers will be searching for evidence of quality, so the author's reputation will be an important cue. This suggests that authors would not want to publish work of low quality. The readership of the paper, readers' comments, and reviews might serve as further useful cues to potential readers.

---

* Papers with findings on peer review appear to be difficult to publish. I would not have attempted to write this current paper had it not been solicited. That said, I sought, and received much pre-submission peer review. The editor also had the paper reviewed by two anonymous referees.

Authors should seek peer reviews before submitting papers for publication. This is an important part of the research process. For example, in preparing this paper, I sent copies to 45 people to seek peer review, and 26 made comments. As an editor and reviewer, I sometimes wish that all researchers relied on pre-submission peer reviews, because it is often painfully obvious that they have not done so. My impressions are consistent with the findings of MacNealy, Speck & Clements (1994) in a survey of authors: For 96 papers published in leading technical communication journals, 44% of the authors reported that they had not solicited opinions from colleagues before submission.

It is distressing that authors often ignore peer review when their paper is rejected by a journal. Wilson (1978) found that 85% of the papers rejected by the *Journal of Clinical Investigation* were eventually published elsewhere, and the majority of these were either not changed or changed in only minor ways. Lock & Smith (1986) obtained similar results for papers rejected by the *British Medical Journal*. Hargens (1990) reached a similar conclusion for papers in the social sciences. Yankauer (1985), in a study of 61 papers that had been rejected by another journal before they were submitted to the *American Journal of Public Health*, found that 48% had not been revised either moderately or substantially. Patterson & Smithey (1990), in a study of papers rejected by the *American Political Science Review*, concluded that of the 263 papers that were then submitted to another journal, 43% contained no revisions based upon the APSR reviews.

EFFECTS OF CHANGES ON QUALITY AND FAIRNESS

The above suggestions were made to increase the likelihood of publishing innovative findings. This section speculates on what the effects might be on quality and fairness.

Some of the above changes put more emphasis on authors to ensure the quality of their paper. Those who do not will have difficulty publishing further papers because editors would be less likely to invite them, and editors and readers would also be wary of the quality of their work.

The current system of journal peer review encourages authors to be concerned primarily about satisfying two or three reviewers. To satisfy these reviewers, authors might have to make changes that harm the paper. In Bradley's (1981) author survey about the acceptance of their last manuscript that required changes, 27% of 348 psychologists and statisticians said that they made reviewers' changes even though they believed they were incorrect. My suggestions are designed to orient authors to consider what might happen with post-publication peer review.

The current practice of giving the responsibility of peer review to two or three reviewers might lead editors, authors, and readers to conclude that once the paper is accepted, its quality is established. I, like others, have been frustrated when asking journals to publish corrections for published papers. The proposed policy changes should facilitate corrections.

If journal peer review led authors to forgo seeking pre-submission peer review, the net effect on quality might be negative. For example, in Bradley (1981), psychologists and statisticians were asked, "In writing your next article, if you knew that instead of being refereed, it would be published as you submitted it (except for changes to conform to the journal's format), how careful would you be in preparing it?"; 21% said more careful, six percent said less, and the remainder said equally careful.

Post-publication review can help to provide a fair assessment of authors for hiring and promotion decisions. Citation indices are particularly helpful for assessing the quality of work some years after its publication. Citations seem to generally be positive; Spiegel-Rosing (1977) analyzed citations to papers in *Science Studies* from 1971 through 1974 and found less than one-half of one percent of the citations were negative. To focus attention on citations, the *Administrative Science Quarterly* offers a prize for authors of its most frequently cited papers over a five-year period.

Some of the changes I have discussed, especially those listed under "alternate formats," would decrease the use of paper counts for evaluating researchers. This might lead to a greater emphasis on measures of impact, such as whether readers believe the findings to be important, whether the paper is read, whether it is cited by other researchers, whether others use the findings, and whether it contains serious errors. I see such criteria as aiding fairness in the evaluation of researchers. A paper that is ignored for ten years, say, might be considered to be of little value. Errors discovered after publication can be linked to the paper for easy access by evaluators. Authors would be motivated to provide papers that were free of errors because they would find it unpleasant to have these errors publicized. Negative citations can alert the scientific community to shortcomings in published work. For example, citations to the fraudulent research by Breuning were mostly negative, and they were increasingly negative over time (Garfield & Welljams-Dorof, 1990).

CONCLUSIONS

Efforts by journals to ensure quality and fairness through peer review have not been overly successful. They also reduce the likelihood that important controversial findings will be published, or at least they delay them. This problem is expected to become worse as the number of submissions increases. Paradoxically, then, the increase in the number of submissions is expected to lead to a decrease in the proportion of published papers with innovative findings. This is consistent with the findings of Holub, Tappeiner & Eberharter (1991) that important papers are decreasing over time as a percentage of published papers. One would expect the most serious losses to occur for the leading journals.

If we want to publish important innovative findings, then we should develop procedures to encourage this. Screening might be changed from a "yes-no" decision to a decision about how to report results. This means deciding how much space should be given, whether the journal should qualify its support by adding a note of caution, and whether the research should be published on paper, electronically, or both.

Electronic publication by journals can address many of the problems associated with peer review, and it helps to solve the major problem in the current system, namely, the shortage of space. It will greatly reduce the cost of disseminating knowledge. This means that instead of screening papers, journals can simply publish the papers along with their judgment as to the worth of the research. Electronic publishing will also improve the ease of conducting post-publication peer review.

To increase the publication of innovative findings, steps can be taken by editors, reviewers, and authors. The rate of advancement of knowledge in science and engineering should increase rapidly as barriers to the publication of innovative research findings are reduced. That said, whenever journals change peer review procedures, I recommend that they make changes gradually and monitor the effects carefully. In addition to assessing changes in the number of innovative findings that are published, the effects on quality and fairness should be assessed.

Acknowledgments: Many people provided useful comments. Among them were Janice M. Beyer, Fred Collopy, William M. Epstein, Richard H. Franke, Raymond Hubbard, Joel Kupfersmid, Byron Sharp, Arthur E. Stamps III, Brian Wansink, David Watkins, and anonymous reviewers. This does not imply that they agreed with my conclusions. Mary Haight and DaraYang provided editorial assistance.

REFERENCES

In the references below, the empirical studies that were used in this paper are designated by an "E". If the paper was based on an experiment or quasi-experiment, an "X" follows the "E". If an author of this paper replied to my request for information about the coding, or if the paper was by the author of the current paper, the paper is designated by an "ER", or "EXR" for experimental papers.

Abramowitz, S. I., Gomes, B. & Abramowitz, C. V. (1975) Publish or politic: Referee bias in manuscript review. *Journal of Applied Social Psychology* **5**, No. 3: 187-200. EX

Abrams, P. A. (1991) The predictive ability of peer review of grant proposals: The case of ecology and the US National Science Foundation. *Social Studies of Science* **21**: 111-132. ER

Arkes, H. (1996) The persistence of management folklore. *Interfaces* **26**, No. 4: 42-44.

Armstrong, J. S. (1980) Unintelligible management research and academic prestige. *Interfaces* **10** (April): 80-86. EXR

Armstrong, J. S. (1982) Barriers to scientific contributions: The author's formula. *The Behavioral and Brain Sciences* **5**: 197-199. ER

Armstrong, J. S. (1985) *Long-Range Forecasting*. John Wiley, New York.

Armstrong, J. S. (1996) Management folklore and management science: On portfolio planning, escalation bias, and such (with commentaries). *Interfaces* **26**, No. 4: 25-55.

Armstrong, J. S. & Hubbard, R. (1991) Does the need for agreement among reviewers inhibit the publication of controversial findings? *Behavioral and Brain Sciences* **14** (March): 136-137. ER

Atkinson, D. R., Furlong, M. J. & Wampold, B. E. (1982) Statistical significance, reviewer evaluations, and the scientific process: Is there a statistically significant relationship? *Journal of Counseling Psychology* **29**, No. 2: 189-194. EX

Bakanic, V., McPhail, C. & Simon, R. J. (1990) If at first you don't succeed: Review procedures for revised and resubmitted manuscripts. *American Sociologist* **21**, No 4: 373-391. E

Barber, B. (1961) Resistance by scientists to scientific discovery. *Science* **134**: 596-602.

Begg, C. B. & Berlin, J. A. (1988) Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society* **A 151**: 419-463. E

Beyer, J. M., Chanove, R. G. & Fox, W. B. (1995) The review process and the fates of manuscripts submitted to AMJ. *Academy of Management Journal* **38**: 1219-1260. ER

Blank, R. M. (1991) The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *American Economic Review* **81**: 1041 1067. EXR

Bradley, J. V. (1981) Pernicious publication practices. *Bulletin of the Psychonomic Society* **18**: 31-34. E

Broad, W. & Wade, N. (1982) *Betrayers of the Truth.* Simon and Schuster, New York.

Campanario, J. M. (1995) On influential books and journal articles initially rejected because of negative referees' evaluations. *Science Communication* **16** (March): 304-325. ER

Chalmers, I. (1990) Underreporting research is scientific misconduct. *Journal of the American Medical Association* **263**, No. 10: 1405-1408.

Chalmers, T. C., Frank, C. S. & Reitman, D. (1990) Minimizing the three stages of publication bias. *Journal of the American Medical Association* **263**, No. 10: 1392-1395. E

Christiansen, D. (1978) The perils of publishing. *IEEE Spectrum* **15**, No. 5: 27.

Cho, M. K. & Bero, L. A. (1996) The quality of drug studies published in symposium proceedings. *Annals of Internal Medicine* **124**, No. 5: 485-489. E

Cicchetti, D. V. (1991) The reliability of peer review for manuscript and grant submissions: A cross disciplinary investigation. *Behavioral and Brain Sciences* **14** (March): 119-186. E

Cohen, J. (1994) The earth is round (p < .05). *American Psychologist* **49**: 997-1003.

Davidson, R. A. (1986) Source of funding and outcome of clinical trials. *Journal of General Internal Medicine* **1**: 155-158. ER

Dewald, W. G., Thursby, J. G. & Anderson, R. G. (1986) Replication in empirical economics: The Journal of Money, Credit, and Banking project. *American Economic Review* **76**: 587 603. E

Eichorn, P. & Yankauer, A. (1987) Do authors check their references? A survey of accuracy of references in three public health journals. *American Journal of Public Health* **77**: 1011-1012. ER

Epstein, W. M. (1990) Confirmational response bias among social work journals. *Science, Technology, and Human Values* **15**: 9-38. EXR

Evans, J. T., Nadjari, H. I. & Burchell, S. A. (1990) Quotational and reference accuracy in surgical journals: A continuing peer review problem. *Journal of the American Medical Association* **263**, No. 10: 1353-1354. E

Fagan, W. T. (1990) To accept or reject: Peer review. *Journal of Educational Thought* **24**: 103-113. ER

Fletcher, R. H. & Fletcher, S. W. (1997) Evidence for the Effectiveness of Peer Review. *Science and Engineering Ethics* **3**: 35-50. EX

Franke, R. H., Edlund, T. W. & Oster, F. (1990) The development of strategic management: Journal quality and article impact. *Strategic Management Journal* **11**: 243-253. ER

Fuess, S. M. (1996) On replication in business and economics research: The QJBE case. *Quarterly Journal of Business and Economics* **35**, No. 2: 3-13. E

Friedman, P. J. (1990) Correcting the literature following fraudulent publication. *Journal of the American Medical Association* **263**: 1416-1419. ER

Gans, J. S. & Shepherd, G. B. (1994) How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives* **8**, No. 1: 165-179. E

Garcia, J. (1981) Tilting at the paper mills of academe. *American Psychologist* **36**, No. 2: 149-158.

Garfield, E. & Welljams-Dorof, A. (1990) The impact of fraudulent research on the scientific literature. *Journal of the American Medical Association* **263**, No. 10: 1424-1426. E

Goodstein, L . & Brazis, K. (1970) Credibility of psychologists: An empirical study. *Psychological Reports* **27**, No. 3: 835-838. E

Gottfredson, S. D. (1978) Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist* **33**: 920-934. EX

Greenwald. A. G. (1975) Consequences of prejudice against the null hypothesis. *Psychological Bulletin* **82**: 1-20. E

Hargens, L. L. (1990) Variation in journal peer review systems: possible causes and consequences. *Journal of the American Medical Association* **263**: 1348-1352.

Harnad, S. (1979) Creative disagreement. *The Sciences* **19**: 18-20.

Hendrick, C. (1976) Editorial comment. *Personality and Social Psychology Bulletin* **2**, No. 3: 207-208.

Hibbitts, B. J. (1996a) Last writes? Re-assessing the law review in the age of cyberspace. http://www.law.pitt.edu/hibbitts/lastrev.htm; version 1.1, June 4, 1996; *New York University Law Review* **17**: 615-688.

Hibbitts, B. J. (1996b) Yesterday once more: Skeptics, scribes and the demise of law reviews. *Akron Law Review* **30** (forthcoming).

Holub, H. W., Tappeiner, G. & Eberharter, V. (1991) The iron law of important articles. *Southern Economic Journal* **58**: 317-328. ER

Horrobin, D. F. (1982) A philosophically faulty concept which is proving disastrous for science. *Behavioral and Brain Sciences* **5**, No. 2: 217-218.

Horrobin, D. F. (1990) The philosophical basis of peer review and the suppression of innovation, *Journal of the American Medical Association* **263** (March 9): 1438-1441.

Horrobin, D. F. (1996) Peer review of research grant application. *Lancet* (forthcoming).

Hubbard, R. & Armstrong, J. S. (1992) Are null results becoming an endangered species in marketing? *Marketing Letters* **3**: 127-136. ER

Hubbard, R. & Armstrong, J. S. (1994) Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing* **11**: 233-248. ER

Hubbard, R. & Vetter, D. (1996) An empirical comparison of published replication research in accounting, economics, finance, management and marketing. *Journal of Business Research* **35**: 153-164. ER

Jauch, L. R. & Wall, J. L. (1989) What they do when they get your manuscript: A survey of Academy of Management reviewer practices. *Academy of Management Journal* **32**: 157-173. E

Juhasz, S., Calvert, E., Jackson, T., Kronick, D. A. & Shipton, J. (1975) Acceptance and rejection of manuscripts. *IEEE Transactions of Professional Communications* **PC18**: 177-184. E

Kerr, S., Tolliver, J. & Petree, D. (1972) Manuscript characteristics which influence acceptance for management and social science journals. *Academy of Management Journal* **20**, No. 1: 132-141. ER

King, D. W., McDonald, D. D. & Roderer, N. K. (1981) *Scientific Journals in the United States: Their Production, Use, and Economics*. Hutchison Ross, Stroudsburg, Pa. E

Koehler, J. J. (1993) The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes* **56**: 28-55. EXR

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.

Kupfersmid, J. & Wonderly, D. M. (1994) *An Author's Guide to Publishing Better Articles in Better Journals in the Behavioral Sciences*. Clinical Psychology Publishing Co., Brandon, Vermont. ER

Lau, R. R. (1994) An analysis of the accuracy of "trial heat" polls during the 1992 presidential election. *Public Opinion Quarterly* **58**: 2-20.

Lindsey, D. (1978) The Scientific Publication System in Social Science. Jossey Bass, San Francisco. E

Liversidge, A. (1989) PNAS publication of AIDS article spurs debate over peer review. *The Scientist* **3**, No. 7: 4-5, 19.

Lock, S. & Smith, J. (1986) Peer review at work. *Scholarly Publishing* **17**, No. 4: 303-316. E

Lock, S. & Smith, J. (1990) What do peer reviewers do? *Journal of the American Medical Association* **263**, No. 10: 1341-1343. E

Lloyd, M. E. (1990) Gender factors in reviewer recommendations for manuscript publication. *Journal of Applied Behavior Analysis* **23**: 539-543. EX

MacNealy, M. S., Speck, B. W. & Clements, N. (1994) Publishing in technical communication journals from the successful author's point of view. *Technical communication* **41**, No. 2: 240-259. E

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

Marsh, H. W. & Ball, S. (1989) The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education* **57**, No. 2: 151-169. E

McCloskey, D. N. & Ziliak, S. T. (1996) The standard error of regressions. *Journal of Economic Literature* **34** (March): 97-114. E

McNutt, R. A., Evans, A. T., Fletcher, R. H. & Fletcher, S. W. (1990) The effects of blinding on the quality of peer review: A randomized trial. *Journal of the American Medical Association* **263**, No. 10: 1371-1376. EX

Munley, P. H., Sharkin, B. & Gelso, C. J. (1988) Reviewer ratings and agreement on manuscripts reviewed for the Journal of Counseling Psychology. *Journal of Counseling Psychology* **35**, No. 2: 198-202. E

Murray, G. D. (1988) The task of a statistical referee. *British Journal of Surgery* **75**: 664-667. E

Needleman, H. L. (1992) Salem comes to the National Institutes of Health: Notes from inside the crucible of scientific integrity. *Pediatrics* **90**, No. 6: 977-981.

Newcombe, R. G. (1987) Towards a reduction in publication bias. *British Medical Journal* **295** (12 September): 656-659.

Patterson, S. C. & Smithey, S. K. (1990) Monitoring scholarly journal publication in political science: The role of the APSR. *PS: Political Science and Politics* **23**: 647-656. ER

Perlman, D. (1982) Reviewer "bias": Do Peters and Ceci protest too much? *The Behavioral and Brain Sciences* **5**: 231-232. E

Pfeifer, M. P. & Snodgrass, G. L. (1990) The continued use of retracted, invalid scientific literature. *Journal of the American Medical Association* **263**, No. 10: 1420-1423. E

Peters, D. P. & Ceci, S. J. (1982) Peer-review practices of psychology journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences* **5**: 187-195. EX

Pressman, S. (1994) Simultaneous multiple journal submissions: The case against. *American Journal of Economics and Sociology* **53**: 316-333.

Rodman, H. & Mancini, J. A .(1977) Errors, manuscripts, and equal treatment. *Research in Higher Education* **7**: 369-374. ER

Roediger, H. L. (1987) The role of journal editors in the scientific process. In: Jackson, D. N. & Rushton, J. P. (eds.), *Scientific Excellence*. Sage Publications, London.

Rowney, J. A. & Zenisek, T. J. (1980) Manuscript characteristics influencing reviewers' decisions. *Canadian Psychology* **21**, 17-21. ER

Rushton, J. P. & Ankney, C. D. (1996) Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin and Review* **3**, No. 1: 21-36.

Salsburg, D. S. (1985) The religion of statistics as practiced in medical journals. *American Statistician* **39**: 220-223. E

Shea, C. (1996) Psychologists debate accuracy of 'significance test. *The Chronicle of Higher Education* **42** (August 16): A12 & A17.

Simon, R., Bakanic, V. & McPhail, C. (1986) Who complains to editors and what happens. *Sociological Inquiry* **56**: 259-271. ER

Smart, R. G. (1964) The importance of negative results in psychological research. *Canadian Psychologist* **5**: 225-232. E

Speck, B. W. (1993) *Publication Peer Review*. Westport, Connecticut.

Spiegel-Rosing, I. (1977) Bibliometric and content analysis. *Social Studies of Science* **7**: 97-113. E

Stamps, A. E. III (1997) Advances in Reer Review Research: An introduction. *Science and Engineering Ethics* **3**: 3-10. ER

Stamps, A. E. III (1997) Using a Dialectical Scientific Brief in Peer Review. *Science and Engineering Ethics* **3**: 85-98.

Sterling, T. D. (1959) Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association*, **54**: 30-34. ER

Sterling, T. D., Rosenbaum, W. L. & Weinkam, J. J. (1995) Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician* **49**: 108-112. ER

Stewart, W. W. & Feder, N. (1987) The integrity of the scientific literature. *Nature* **325**: 207-214. ER

Szenberg, M. (1994) Disseminating scholarly output: The case for eliminating the exclusivity of journal submissions. *American Journal of Economics and Sociology* **53**: 303-315.

Taubes, G. (1994) Peer review in cyberspace. *Science* **266**: 967.
Tetlock, P. E. & Kim, J. I. (1987) Accountability and judgment process in a personality prediction task. *Journal of Personality and Social Psychology* **52**: 700-709.
Wade, N. (1976) IQ and heredity: Suspicion of fraud beclouds classic experiment. *Science* **194**: 916-919.
Weiss, D. J. (1989) An experiment in publication: Advance publication review. *Applied Psychological Measurement* **13**: 1-7.
Wilson, J. D. (1978) Peer review and publication. *Journal of Clinical Investigation* 61, 1697-1701.
Yankauer, A. (1985) Peering at peer review. *CBE Views* **8**, No. 2: 7-10. ER
Yankauer, A. (1990) Who are the peer reviewers and how much do they review? *Journal of the American Medical Association* **263**: 1338-1340. ER