Conceptual Debates and Empirical Evidence about the Peer Review Process for Scholarly Journals

Sandra Thomas
CONCEPTUAL DEBATES AND EMPIRICAL EVIDENCE ABOUT THE PEER REVIEW PROCESS FOR SCHOLARLY JOURNALS

SANDRA P. THOMAS, PhD, RN, FAAN

Until recently, nursing scholars have seldom questioned the underlying premises of peer review, nor have they engaged in the conceptual debates about the peer review process. In this article, literature from a wide range of nursing and nonnursing journals was examined to provide an overview of (a) the conceptual debates and (b) the empirical evidence about the peer review process in scholarly journals. A multiplicity of questions for future research are proposed. (Index words: Scholarly journals; Peer review) J Prof Nurs 27:168–173, 2011. © 2011 Elsevier Inc. All rights reserved.

Peer review arouses very diverse emotions, beliefs, and ambitions. It angers, it reassures, it intimidates, it tramples egos, and it puffs them up. For some, peer review demonstrates the vacuousness and unreliability of social science; for others, the substance and reliability of social science. (Starbuck, 2003, p. 348)

Designed to guard against errors and ensure the quality of published work, peer review has been a hallowed tradition of the scientific community since the 1600s, when the first learned societies were formed (Bjorgo, 2003). At first, a “peer” was simply a member of one of the early European academies such as the Académie Royale des Sciences de Paris. In England, a “Committee on Papers” was established by the Royal Society of London more than 300 years ago to scrutinize manuscripts before they could be published (Kronick, 1990). Before the development of scientific journals, peer reviewers had a rather light workload, simply reviewing book manuscripts from time to time. They would no doubt be astonished at the magnitude of the task now, when scientists are producing thousands of papers every day, requiring thousands of their peers to conduct reviews (Weller, 2001). Contemporary reviewers are still asked to assess the quality of submissions, as well as other aspects, for example, the importance of the topic, contribution to the body of knowledge about the topic, relevance to journal readers, soundness of methods, ethics, and completeness of the report (Jefferson, Rudin, Brodkey Folse, & Davidoff, 2007). It is not clear, however, whether peer review achieves these aims or falls short, even producing harmful effects in some instances. Some of the underlying premises of peer review are seldom questioned within the community of scholars. The purpose of this article was to provide an overview of the conceptual debates about the contemporary peer review process, along with empirical evidence regarding its strengths and weaknesses. Literature from a wide range of nursing and nonnursing journals was examined.

Conceptual Debates About Peer Review

Horrobin (1982), one of the harshest critics of peer review, claims that it is a faulty concept based on two myths: (a) all scientists are peers of one another, and (b) genius will be recognized. In regard to the first myth, there are several author questions, such as “Who are my peers?” and “Why should there be a hierarchical, even adversarial, relationship between my ‘peer’ reviewers and me?” A 2003 survey of authors of published management science papers found that 34% reported being treated like inferiors by journal editors or reviewers, and 56% felt that editors regarded a reviewer’s knowledge of the author’s research as more important than the author’s own knowledge (Bedeian, 2003). Furthermore, 38.7% of the respondents believed that

Professor and Chair, PhD Program in Nursing, University of Tennessee, Knoxville, College of Nursing, Knoxville, TN 37996-4180.
Address correspondence to Dr. Thomas: University of Tennessee, Knoxville, College of Nursing, 1200 Volunteer Blvd., Knoxville, TN 37996-4180. E-mail: sthomas@utk.edu
8755-7223/10/8 - see front matter

168

requested revisions were based on the personal preferences of editors or reviewers.

With regard to the second myth, Campanario (1995) has noted that authors who stray too far from conventional theories are likely to encounter negative reviewer evaluations and rejections. According to Epstein (1995, p. 884), many authors believe that "the more innovative and important their research, the greater the likelihood that their articles would be rejected." Consistent with this perception, Campanario's research found that papers about eight Nobel-Prize-winning discoveries were initially rejected. Nobel recipient Yalow (1982, p. 60) cited her rejection letter in her Nobel lecture, observing that "The truly imaginative are not being judged by their peers. They have none!" One maverick scientist deliberately chose nonreviewed journals for publication of his work because he believed that referees would demand certain changes; despite this nontraditional path, he received the Nobel Prize (Frey, 2003).

Other scientists, aware that university promotion and tenure committees require publication in refereed journals, cannot afford to take a nontraditional path. Critics of the current peer review system allege that as a result, authors are forced to "market" their work, showing that it makes an original (but not too original) contribution, then capitulating to reviewer and editor demands and persisting through the tedious revision process (Starbuck, 2003). Given the high rejection rates of many prestigious journals, authors may simply be grateful for the opportunity to revise.

Frey (2003) proffered a third myth: that reviewers act in the interest of science as a whole, or at least in accord with the interests of the journals they serve. In regard to this myth, Frey and others (e.g., Kumashiro, 2005) argue that reviewers cannot be completely objective; they cannot completely set aside their theoretical perspectives, academic training, and personal opinions. For example, a nurse reviewer who believes that all nursing research should be framed within nursing theories is not inclined to favorably rate a study framed in a theory "borrowed" from psychology. Thus, Bedeian (2004) calls peer review a social process that influences how knowledge claims are presented and validated. Tensions among authors, editors, and reviewers are inevitable, with authors vigorously making their knowledge claims, reviewers rebutting or rejecting them, and editors seeking to balance the tensions (Bedeian, 2004).

Even the purposes of peer review continue to be debated. Perhaps with the aim of maintaining a journal's high rejection rate, some reviewers believe that it is their responsibility to detect limitations in a manuscript (Van Lange, 1999). Social psychologist Van Lange (1999, p. 2550) has created an acronym, "the SLAM effect in peer review," to describe a tendency among some reviewers to Stress Limiting Aspects of Manuscripts. In accord with this stance, Epstein (1995, p. 884) alleged that reviewers frequently exhibit an "I gotcha" mentality, seeming to delight in the discovery of relatively minor limitations that would justify a recommendation of "reject."

Others think that peer review should serve a developmental function for fledgling authors. Nurse editors, in particular, usually encourage reviewers to provide authors with positive feedback as well as negative and to make constructive suggestions for revision of their papers. The intimidation of a novice author about the first few submissions to scholarly journals is well known. Profound demoralization can occur when reviewers are highly critical of their work.

However, there is little agreement about the degree of mentoring that journal reviewers and editors should provide. Eysenck (2000) once stated, rather emphatically, that referees must not act as "nannies."

As a final topic for further debate, Jefferson, Wager, and Davidoff (2002) argued that the objectives of peer review have not yet been clearly identified, making it impossible to assess whether they are being met. What are the objectives? Are reviewers gatekeepers or facilitators?

Strengths of the Peer Review Process

Several studies have identified strengths of the peer review process. In a survey of reviewers for a medical education journal, Snell and Spencer (2005) found that the reviewers took their responsibility seriously, spending considerable time (M = 3 hours) preparing their comments. Kearney, Baggs, Broome, Dougherty, and Freda (2008) found that nursing journal reviewers (N = 1,675) spent even more time reviewing manuscripts—on average, they spent approximately 5 hours on each review, and they reviewed seven to eight manuscripts per year. Reviews (N = 464) conducted by reviewers for Nursing Research were judged as "balanced and fair" in a recent study by Henly and Dougherty (2009, p. 25), although review quality was better for the methods section of the manuscripts than for the theory or literature review sections.

Authors in Weller's (1996) study have acknowledged that peer reviewers helped them to make substantial improvements to their manuscripts, and Jefferson et al. (2007) concluded that peer review contributed to manuscript readability and improved quality of reporting. An international survey of 3,040 academics commissioned by the Publishing Research Consortium (Ware, 2008) found that both authors and reviewers held quite positive views of peer review. For example, 90% of researchers said that peer review had improved their last published articles, 85% agreed that peer review helped scientific communication, and 83% believed that without peer review, there would be no control. Opinions of authors appear to vary, however, based on the eventual disposition of their manuscripts. In a study by Weber, Katz, Waeckerle, and Callahan (2002), authors whose papers were accepted were much more satisfied with their reviews than were those whose papers had been rejected. Satisfaction of authors with the review process was only modest, overall (rated 3 on a 5-point scale).
Flaws of the Peer Review Process

Perhaps the most frequently cited flaw of the peer review process is its susceptibility to bias, that is, reviewer judgments based on factors other than scientific merit of the work (Starbuck, 2003). Myers (1990, p. 64), for example, said: "Almost every scientific researcher I have interviewed has an anecdote about a referee who reviewed an article of his [sic] unfairly, or who requested alterations that, in the writer's view, diminished the value of the article." There is empirical evidence of gender bias, language bias, specialty bias, a general bias toward publishing positive findings, and nationality bias (Ware, 2008). To circumvent reviewer bias, most nursing journals follow the practice of double-blind reviewing (i.e., neither author nor reviewer knows the identity of the other; Baggs, Broome, Dougherty, Freda, & Kearney, 2008). Studies have shown that both nurse reviewers and medical reviewers believe that blinding does decrease bias and encourages fair assessments (Baggs et al., 2008; Regehr & Bordage, 2006).

Nevertheless, even if a reviewer is blinded to the author's identity, there are other aspects of the manuscript that could evoke reviewer bias: the topic, writing style, vocabulary, and/or use of culturally specific idioms or metaphors, to name a few. Prejudice and stereotyping often exist at a preconscious or unconscious level (Plous, 2003). The findings of a study could threaten, irritate, or disturb a reviewer. Kumashiro (2005) argues that the peer review process excludes voices from the margins, citing scathing reviews he himself has received—often angry in tone and dishonest in style or the troubling questions he raised about the mainstream status quo. One editor participating in a survey by Freda and Kearney (2005a) proposed that reviewers sign forms indicating no conflict of interest before reviewing manuscripts, but this has not received wide attention in the literature.

Another flaw of the current system is that peer reviews are inconsistent (e.g., correlated only .12 in Starbuck, 2003, and agreeing no more often than would be expected by chance in Rothwell & Martyn, 2000). This inconsistency among reviews leaves authors puzzled about how to meet the demands of multiple divergent reviewers when they prepare their revisions. As a journal editor, I can attest to the relatively high frequency of the "split review," wherein one reviewer enthusiastically recommends acceptance of the manuscript and the other just as strongly recommends "rejection," necessitating recruitment of at least one more reviewer before I make the final decision about manuscript disposition. Some editors recruit as many as six reviewers if conflict about a manuscript is anticipated. Understandably, an author could feel overwhelmed when receiving six reviews, especially if there is considerable inconsistency among the critiques and inadequate guidance from the editor about which comments deserve emphasis when preparing the revisions.

A classic study illuminating the unreliability of reviews was conducted by Peters and Ceci (1982). In their study, 12 recently published articles were resubmitted to the journals that had just published them. Editors or reviewers detected the ruse for 3 articles, but 8 of the remaining 9 were rejected, allegedly because of serious flaws. In another study, eight errors were inserted in a manuscript, which was then reviewed by 221 respondents; on average, only two errors were detected by the reviewers (Godlee, Gale, & Martyn, 1998).

Several observers (e.g., Bedian, 2003; Spector, 1998) note that editor and reviewer comments have become much more lengthy and detailed in recent years, requiring not only substantial revision of the manuscript itself but also preparation of a veritable treatise to accompany the resubmission. Frey (2003, p. 208), for example, said that "Often, an almost completely new paper is demanded." After completing one revision, it is not uncommon for an author to be asked to revise the paper again; in discussions at professional meetings, I have heard of as many as four required revisions. The expenditure of that much time and effort on a single paper would seem to constitute excessive cost in relation to the ultimate benefit of the paper's publication.

The length of the review process has also generated increasing complaints. Ellison (2002) pointed out that the process took longer in 2000 than in 1970, at least in the field of economics. Indeed, according to Frey (2003), 1 year is not unusual in economics. However, there is considerable variability among disciplines. Ware (2008) reported that reviewing times were longest in humanities and social sciences journals and shortest in medical and nursing journals. Gross and Fonteyn (2003, p. 87) have noted that reviews in nursing journals might take "up to six months," but some medical journals expect reviews to be completed within 2 to 3 weeks (Sellke, 2003). Nurse editors surveyed by Freda and Kearney (2005b) voiced frustration regarding the timeliness of reviewer responses. Although a high percentage (62%) of authors were satisfied with peer review times in Ware's (2008) survey, reviewers perceived themselves to be overburdened. The large majority of reviewing (79%) was done by a core group of active reviewers, whereas other reviewers conducted reviews only sporadically.

A number of authors have voiced alarm about the negativity of the peer review process; Miner (2003, p. 339), for example, expressed "deep concern about what the [review] process is doing to the egos of authors and prospective authors." Van Lange (1999) asserted that reviewers' appraisals are more negative than those of editors, perhaps because reviewers have the protection of anonymity and editors do not. Kumashiro (2005, p. 261) argues that even a constructive review can be problematic: "Strengthening a manuscript begs the question of what we mean by stronger and who determines." The title of the book How to Survive Peer Review (Wager, Godlee, & Jefferson, 2002) is eloquent testimony to the stressful nature of the review process for many authors;
presumably, such a book would not be needed if the process were less arduous.

Some critics of the current peer review system even speak of intellectual prostitution, gamesmanship, and compromise. For example, Frey (2003, p. 205) has alleged that “authors only get their papers accepted if they intellectually prostitute themselves by slavishly following the demands made by anonymous referees.” According to Frey, this prostitution is necessary because of the high journal rejection rates and the extreme time pressure experienced by tenure-track junior faculty: “either they are able to publish in a ‘core’ refereed journal, or they have to bury their dream of an academic career” (p. 210). Fighting the demands made by reviewers is viewed by Frey as a very risky strategy with little possibility of success.

There is some evidence of author gamesmanship. Indeed, Kumashiro (2005) conducted workshops for the explicit purpose of teaching scholars of color “how to play the game.” Twenty percent of the respondents to Bedelian’s (2003) survey said that they had included a reference primarily because they hoped the author of that paper would be chosen to review their manuscript, and almost a quarter of the respondents admitted that they actually made changes they felt were incorrect based on reviews. Choosing your own reviewers, when a journal offers that option, can be another way for authors to “play the game.” Summarizing several studies presented at the Fifth International Congress on Peer Review and Biomedical Publication, Grimm (2005) reported evidence that author-suggested reviewers were more likely to recommend manuscript acceptance. Some journals also permit authors to exclude reviewers, and one study cited by Grimm showed that the odds of acceptance were twice as high when authors had excluded certain reviewers: “People know their assassins,” commented the researcher (Goldsmith, cited in Grimm, 2005, p. 1974).

The solution of many authors when faced with reviewer demands appears to be compromise. This does not necessarily set well with authors, however, as noted by Roth (2002, p. 215):

When one of my articles is finally published, I always have a sense that I am only partially the author, something is lost; this something may well be a part of myself. There are so many other actors...who have...succeeded in making changes...or getting me to make it “in a satisfactory way” that it no longer feels like mine in the traditional sense.

**Implications for the Discipline**

In this paper, I have introduced readers to the conceptual debates and some empirical evidence about the contemporary peer review process for scholarly journals. It is clear that “peer review is imperfect,” as succinctly stated by Weber et al. (2002, p. 2793). Unfortunately, it is only within the past decade that the discipline of nursing has become engaged in the debates about peer review and in conducting empirical research on the process. Although six large international conferences on peer review have been held, only during the sixth conference (2009) were any nurses included as presenters (Broome, Dougherty, Baggs, Kearney, & Freda, 2009; Dougherty, Freda, Kearney, Baggs, & Broome, 2009; Shattell, Chinn, Thomas, & Cowling, 2009).

A five-editor research team that includes Baggs, Broome, Dougherty, Freda, and Kearney has contributed much useful information about reviewer perceptions of peer review because their survey yielded such a large sample (1,675 reviewers from 44 countries) and elicited data on a variety of issues (Baggs et al., 2008; Freda et al., 2009; Kearney et al., 2008). Like reviewers in the study by Snell and Spencer (2005), nurse reviewers studied by Kearney et al. (2003) were motivated to participate in the peer review process because they felt it afforded the opportunity to keep up to date with the newest ideas in the field, as well as contributing to the advancement of science (Kearney et al., 2008).

Study findings by the five-editor team that should arouse concern in the discipline are as follows: (a) the minimal training received by most reviewers (65% wanted—but did not receive—formal training), (b) the number of reviews it takes for reviewers to become comfortable with their responsibilities (55% were comfortable after one to five reviews, but 45% needed to do more reviews than that to achieve comfort), and (c) the lack of feedback from editors (87% of reviewers wanted more feedback; Freda et al., 2009). In an earlier study, nurse editors expressed concern about reviewer preoccupation with copyediting (correcting grammar, spelling, or typographical errors); instead of copyediting, editors wanted reviewers to place more emphasis on methods, the comprehensiveness of the literature review, clinical relevance, and other, more substantive aspects of manuscripts (Kearney & Freda, 2005). Henly and Dougherty (2009) were disturbed by the high frequency of poor or inadequate reviews when they scrutinized the quality of reviews for Nursing Research. The need for better reviewer training is apparent, and reviewers want it, but previous research on interventions such as workshops and training manuals provides little evidence of effectiveness (Freda et al., 2009). Perhaps, every new reviewer should complete several “trial reviews” that are critiqued by a mentor before retention on a journal review panel.

Freda et al. (2009) suggested allowing all reviewers of a manuscript to read each others’ reviews, after the editor’s decision about disposition has been made. Many questions are yet to be investigated by nurse researchers. The following questions—not an exhaustive list—are worthy of attention: Is it useful for authors to have the opportunity to suggest inclusion or exclusion of potential reviewers? Is it useful for authors to rate the quality of reviews they receive? What proportion of nurse authors simply withdraw their manuscripts rather than make revisions (and further revisions of the revisions) that they are not comfortable making? Would an appeals
process for improperly reviewed and rejected articles minimize author discontent? If reviewers sign conflict of interest forms, does the prevalence of bias diminish? If reviewers’ identities are revealed to authors, what outcomes ensue? Is there an optimal way to select and train reviewers? What is the minimum number of “trial reviews” that should be required for retaining a potential reviewer on the panel of a journal? Once a journal adopts a procedure for allowing reviewers to read one another’s reviews, is there improvement in reviewer satisfaction and/or review quality? Given a reviewer’s propensity for late reviews, poor review quality, or tendency to make pejorative comments, at what point should an editor remove such a reviewer from the panel? The discipline of nursing needs the answers to these questions.

It is imperative that continued efforts be made to improve the peer review process for nursing’s scholarly journals.

References


Shattell, M., Chinn, P., Thomas, S. P., & Cowling, W., (2009, September). Quality of peer reviews in 3 nursing journals from the perspective of authors and editors. Presented at the Sixth International Congress on Peer Review and Biomedical Publication, Vancouver, BC, Canada.


