Journal Peer Review

Comparing the perceptions and quality judgments of experienced Australian reviewers in Education and the Physical Sciences

Yanping Lu, BA, MLMED

Submitting for the degree of Doctor of Philosophy

October 2010

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

ACKNOWLEDGEMENT

I have spent many wonderful years at The University of Newcastle. This extremely invaluable experience and the completion of this thesis have been made possible with the support I have received from the University, the School of Education, and numerous outstanding individuals.

First and foremost I feel deeply indebted to my supervisor, Professor Allyson Holbrook of The University of Newcastle, whose patience, kindness, thoroughness, and incredible passion and experience in research, have been the ultimate motivation and driving power in my journey through doctoral training. Professor Holbrook has also provided me with the perfect role model of an intelligent, diligent, and creative female academic.

I have been a student member of the Centre for the Study of Research Training and Impact (SORTI) of which my supervisor is the director. SORTI brings together a team of outstanding members who have generously offered me assistance. I owe my deepest gratitude to Dr Ray McCredie Barker (1920-2009) who generously provided a scholarship to support my project. I would also like to thank Professor Sid Bourke who provided statistical advice, Mrs Hedy Fairbairn who assisted in proofreading of the thesis, and other members who offered valuable comments to my progress presentations.

The journey has been made more comfortable and achievable with the support of the School of Education, including an encouraging culture of research, a well-equipped research office dedicated to my exclusive use, two Travel Scholarships that allowed me to experience Melbourne, Australia and Budapest, Hungary, as well as numerous working opportunities in research and teaching. I would also like to thank Associate Professor James Ladwig and Associate Professor Michael Arthur-Kelly, among others, for allowing me to trial the research instruments with them and providing suggestions for improvement.

This thesis would not have been possible without the contribution of the participants to the study, who are busy academics but have been extremely generous in sharing their time and opinions with me.

I have been undertaking a full time job at Planning, Quality and Reporting, The University of Newcastle since 2009. I am indebted to my supervisor at work, Associate Professor Anne Young, and my colleagues for their support of me in the completion of my thesis.

Finally, I would like to express special thanks to my husband Yuhua (Eric) Shi. Eric is also a doctoral candidate and full time employee at the University, and we have a four-year old daughter. He has helped me wherever he could to concentrate on completing this thesis and supported me mentally during the course of this work. Without his love, help and encouragement, this study would not have been completed.

Table of Contents

Table of Contents

Table of Contents

Table of Contents

List of Tables

List of Tables

List of Figures

List of Appendices

CHAPTER 1: RESEARCH ASSESSMENT AND PEER REVIEW: AN OVERVIEW

1.0 Introduction

Peer review is a core activity in academic life and as a means of identifying high quality research plays a major role in determining the effectiveness and direction of research across disciplines. As will be demonstrated in this and subsequent chapters, peer review is a phenomenon of considerable complexity. For instance, Chubin and Hackett (1990, p. 2) define peer review as 'an organized method for evaluating scientific work which is used by scientists to certify the correctness of procedures, establish the plausibility of results, and allocate scarce resources', while McPeek et al. (2009, p. E155) describe peer review as 'an amazing cooperative network'. Peer review is a social system of evaluation and decision making, which involves both established, objective procedures and subjective judgments.

While peer review is used in a number of different settings, this study focuses on journal peer review in particular, with an emphasis on the discipline areas of education, physics and chemistry. Although peer review has attracted substantial empirical attention, most studies focus on isolated aspects and isolated contexts (namely, single journal or discipline). This study attempts a more holistic approach to gain depth of insight into the range of functions, expectations, and activities associated with journal peer review across and within disciplines.

When the study was initiated in 2002, there was an absence of a nationally accepted framework for research assessment, or for ranking scholarly journals in Australia. As these developments were on the horizon (Department of Education, Training and Youth Affairs, 2000), a study of Australian academics' understanding of judging research quality, and attitude toward peer review appeared to be very timely. Also, Australian research in the field is sparse.

The study is reported in nine chapters. The first three chapters identify the background to the study and position the research questions and design within the multi-disciplinary framework of studies of journal peer review. There is also a

substantial published literature that takes the form of 'editorials'. This unexamined body of work provides a rich source of information concerning the debates and issues surrounding peer review and is investigated in the fourth chapter. The remaining chapters present and then draw together the findings and interpretations based on data collected by survey, interview, and review reports for the selected disciplines of education, physics and chemistry.

The current chapter situates the study in a number of ways. It begins with a brief examination of the changing environment and the expectations of 'quality' in the higher education sector, and identifies the main mechanisms adopted in the sector for the purpose of quality assurance and research assessment, including benchmarking, external audit and peer review. Noting the accepted value and role of peer review in this context, it then proceeds to establish the multiple functions of journal peer review. The final section draws together key themes in empirical studies on the context and management of journal peer review. This sets the scene for an examination of the practices of journal peer review in the next Chapter.

It is the individual practice of reviewing, and in particular distinctions in judgment of quality in different disciplines that is the main focus of this thesis. The author also examines very closely the published policies and instructions given to reviewers.

1.1 The Changing Environment and the Expectation of Quality

In the past several decades, the higher education sector worldwide has undergone remarkable changes politically, financially and academically (Armstrong, 1999; Cullen et al., 2003; Lund & Jackson, 2000; Marginson & van der Wende, 2007; Menon, 2003; Reindl, 2004; Smith, 2003; Weeks, 2000; Woodhouse, 2003). The small, elite, highly autonomous higher education 'community' of the 1960s has evolved into a mass system and universities have steered policy directions in response to increasing external pressures (Lund & Jackson, 2000). What has arisen from the changing environment is the quality assurance movement with intensifying attention to accountability in the quality of performance, and the consolidation of external quality assessment agencies and mechanisms.

This in turn has placed 'quality' in the spotlight. Harvey and Green (1993, p. 10) posed the challenge 'What the hell is quality?' In fact the notion of 'quality' as a general term and what comprises quality especially in assessing research has long been the subject of scholarly debate.

Quality deals with a number of complex and often intangible notions (Lindsay, 1994), which makes it easier to recognize than to describe and measure (Harvey & Green, 1993; Relman, 1990; Stephenson, 2004). In the era of mass education, quality has become a 'central term in the lexicon of contemporary education and a major point of interest for various interest groups' (Newton, 2002, p. 44). The groups, including governments, universities, academics, students, and the general community, can perceive quality very differently. Their perceptions, as Harvey and Green (1993, p. 10) noted, are not different notions on the same thing but 'different notions on different things with the same label.' Thus quality can only be viewed as a relative concept – relative to its user and the context in which it is used, and to the process or outcome by which it is judged.

Nebulous as the concept of quality appears to be, some definitions have been generated to advance the discussion of quality. Harvey and Green (1993) grouped the work of earlier theoreticians (for example, Ball, 1985; Kogan, 1986) into five considerations, namely: quality as exceptional or excellence, as perfection or consistency, as fitness for purpose, as value for money, and as transformation (Harvey & Green, 1993, p. 10).

The first three considerations suggest 'levels' of quality. 'Excellence' views quality as exceeding high standards or doing the right things well; quality is judged in terms of meeting standards; and quality is improved if standards are raised. Related to this is the view of 'threshold standards', where the label 'quality' is attributed to all of the products or activities being judged that have met predetermined standards. In terms of 'perfection or consistency', quality is defined as 'zero errors' or consistent flawless outcome. Quality as 'fitness for purpose' views the quality of an output to be meaningful when it fits its pre-defined purpose, typically presented through a mission statement established by an institution for itself. Quality as 'value for money'

focuses on the capability of meeting high standards at reasonable cost, at the core of which is the notion of accountability. Quality as 'transformation' reflects the philosophical notion of 'fundamental qualitative change' of form (Harvey & Green, 1993, pp. 10-25). The thesis will refer back to these considerations of quality at various points.

1.2 Contemporary Quality Assessment Mechanisms in Higher Education

Most quality assessment systems depend on a combination of various mechanisms to assess quality of different kinds (Harman & Meek, 2000). This section examines the current quality assessment mechanisms in higher education, with an emphasis on the situation in Australia. Lund and Jackson (2000) in their edited book, Benchmarking for Higher Education, reviewed the development of contemporary quality assessment systems in the UK and identified a number of mechanisms being adopted for quality assurance purpose, including: benchmarking, external audit, peer review, institutional quality profile, league table, balanced scorecard, self-assessment, and performance indicators. The first three are used very widely, and league tables have more recently gained prominence (Marginson & van der Wende, 2007).

Benchmarking is a practice of self-evaluation and improvement through systematic comparison of performance with competitors, with the purpose to identify strengths and weaknesses and learn ways of adapting and improving when conditions change (Camp, 1989). This practice was originated in Japan and was introduced to the West in the mid 1980s, and then widely adopted by the commercial sector as part of the Total Quality Management movement (Zairi, 1994).

Its introduction to the higher education sector occurred during the quality assurance movement in the 1990s (Harman, 1998), where the transformation of university governance from a collegial system to one emphasising corporate managerialism and internationalism ignited the need for a national benchmarking of university performance (Murphy, 1995; Weeks, 2000). Universities in the US adopted this practice in the early 1990s (Alstete, 1995), followed closely by Australia and the UK (Jackson, 2001). In Australia, the Committee for Quality Assurance in Higher

Education was established in 1992, with the mission to advise the government on the conditions of quality in higher education. It was believed to have kindled the interest in national benchmarking in the sector. Many universities have since joined benchmarking projects hosted by Australian or international tertiary networks such as the Commonwealth Higher Education Management Service Club (Garlick & Pryor, 2004).

In 2000, the Department of Education, Training and Youth Affairs of Australia launched Benchmarking: A manual for Australian universities (McKinnon et al., 2000). The Manual identified 67 benchmarks of nine categories regarding different aspects of university operation. There were 12 research-related benchmarks, among which research impact and outcome were prioritized; however, only the quantity, not quality, of the research outcome was addressed (McKinnon et al., 2000, pp. 103-111).

The other major quality assurance mechanism, external audit, measures 'fitness for purpose'. The strengths of audit include that it is flexible and receptive to institutional diversity; it can often raise issues which may not appear evident to an internal group; and it ensures accountability of outcomes by producing reports independently (Woodhouse, 2003). Its usage in higher education was initiated by the Academic Audit Unit of the Committee of Vice-Chancellors and Principals in the UK in the early 1980s (Lund & Jackson, 2000). External audit agencies were then established in the US, New Zealand, and many European countries.

In the late 1990s, in the pursuit of a suitable model of quality audit for Australian universities, a number of overseas models were consulted, and the New Zealand Universities Academic Audit Unit was used as a model (Vidovich, 2002). In 2000, the Ministerial Council for Employment, Education, Training and Youth Affairs launched the Australian University Quality Agency (AUQA) (Woodhouse, 2002) (now known as Tertiary Education Quality and Standards Agency or TEQSA).

The AUQA audits adopt a 'whole-of-institution' approach, looking more at broad issues than detailed achievements of individual faculties or academics, and reviews process rather than outcomes (Woodhouse, 2002). As holistic approaches, both audit

and benchmarking are largely incapable of assessing specifics, especially those related to research performance. This is where peer review becomes important.

As a mechanism for quality control in research assessment, peer review has salience for a wide range of stakeholders not only funding bodies, universities, or journals, but also individual researchers. It is an autonomous academic process (Vidovich, 2002), and research is believed to '[benefit] from the existence of nationally accepted… practices' (Martin, 2003, p. 18). It is long-standing, globally-accepted and 'as ubiquitous as the air we breathe' (Suls & Martin 2009, p. 40). It offers an approach for validating claims about results, which attends to both macro and micro standards of the research (Cain, 1999, p. 536), and is 'an essential and integral part of consensus building' that is ' inherent and necessary to the growth of scientific knowledge' (Kronick, 1990, p. 1321).

In Australia, the largest national agency that awards competition-based research funding – the Australian Research Council (ARC) – employs a peer review system to assess grant applications. Journal and grant peer review adopt very similar practices of using external peers to evaluate research, yet historically these two systems have developed independently of each other and in different ways (Burnham, 1990; Rennie, 1999). Burnham notes that journal peer review 'appeared sporadically', and was 'influenced by independent evolutionary processes at each journal' (p. 1328). Its operation can be discipline-, journal-, or editor-specified. This study features questions that compare ARC and journal peer review practices (see Chapter 5)

Irregularity in the development of journal peer review underlines the fact that it is a unique phenomenon with unique problems that can be masked by its iconic status. Many have pointed out that the process is beset by dilemmas and uncertainties (Eisenhart, 2002; Emden, 1996; Finn, 2002; Knoll, 1990; Macnab & Thomas, 2007; Smith, 2006a).

The following two sections and the next chapter discuss the focus of debates and the scope of research on journal peer review – referred to hereafter as the peer review literature. As peer review concerns basically all disciplines that have journal outlets, the discussion covers literature collected from all fields, so as to demonstrate

both trends and perspectives, and commonalities and differences, across disciplines. The literature is both empirical and non-empirical and the next section, which essentially begins the literature review, is built upon the non-empirical side of the literature, where insights into the functions of peer review are most clearly articulated.

1.3 Journal Peer Review: An introduction to the debate and dilemmas

Journal peer review has received considerable attention in the past decades and, in the current environment of quality assurance and global research competition, the scrutiny is intensifying. Even though peer review is perceived as the best option in the judgement of research quality, there has been 'a growing tide of criticism [that] has questioned the impartiality, reliability and validity of peer review as a means for determining scientific merit' (Bedeian, 2004, p. 199). Despite an extensive literature on peer review, much of what occurs in practice is still seen to resemble a 'black-box' (Atkinson, 1994, p.155).

The researcher came to the study from an education background and an interest in quality and assessment of quality in education research. She was faced with a diverse and unwieldy literature. Initially, some 1200 articles were identified by a search of online library databases, issues of the Journal of American Medical Association that published articles presented at the five International Congresses on Editorial Peer Review in Biomedical Publication, bibliographies of monographs devoted to the topic, for example, Peer Review: Its strengths and weaknesses (Weller, 2001) and Publication Peer Review: An annotated bibliography (Speck, 1993), and the reference lists of retrieved articles.

The articles accessed were published between 1960 and 2008, including 385 articles reporting original research, some 600 non-empirical articles (for example, essays and commentaries), 176 editorials, and 20 books which focused strictly or substantially on the topic of journal peer review. A database was developed to accommodate the magnitude of the material and to support categorization by type of article, theme, publication date, and discipline.

1.3.1 The concept and definition of peer review

The basic concept of journal peer review, that external experts provide advice to assist editors in the selection of papers for publication, was established many decades ago. Roediger (1987) in Scientific Excellence: Origins and assessment maintained that the first official editorial board was established in the early 1800s for the French Journal des Scavans.

Others argued that the precedent was set by the Royal Society of London when it took over official responsibility for the publication of the Philosophical Transactions in 1752 and established a 'Committee on Papers' (Kronick, 1990; Rennie, 1999; Wilson, 2002). Its aim was to review all papers to be published in the Philosophical Transactions and consult 'any other members of the Society who are knowing and well skilled in that particular branch of Science that shall happen to be the subject matter of any paper which shall be then to come under their deliberations' (cited in Kronich, 1990, p. 1321).

The concept of 'expert judgement' has been so deeply embedded in academic culture that rarely has anybody probed further into what it means. Authors tend to write about peer review without defining it. As for those who offer a definition, they usually do so by describing the procedures they believe peer review should follow. For example, Dancik (1991) once asked, 'What, exactly, is peer review?' and then provided an answer by describing some of its purposes such as 'to identify flaws in design and analysis or interpretation, to suggest improvements, to direct papers to the most appropriate outlets, to discourage repetition in publishing, and to weed out poor science or scholarship' (pp. 91-92). Relman (1990) suggested that 'we need to have a clear understanding of exactly what peer review is' (p. 520) and then he addressed this by describing the editorial procedures of the journal he edited.

The tendency of 'by-passing' the substantive meaning of peer review is a reflection of the restricted debate that has 'focused on everything but its most important aspect – the cognitive task of the reviewer assessing a manuscript' (Kassirer & Campion, 1994, p. 96).

1.3.2 The notion of 'peer'

A dictionary definition of peer is 'a person who is equal in ability, standing, rank, or value' (The Australian Concise Oxford Dictionary, p. 1037). In the context of journal reviewing, a 'peer' is one who is on a par with the author in expertise and academic status. In reality, however, this is often not the case.

Atkinson (1994, p. 155) sees peer review as a 'black-box' undertaken by 'an elite group operating behind a screen of anonymity'. According to Osborne and Brady (2002, p. 165), peers are a group of 'either consensus- or editor-defined' members of the scientific community. It is not surprising therefore that some scholars comment that the term 'peer' in reviewing is a 'misnomer' (Atkinson, 1994; Chubin & Hackett, 1990; Osborne & Brady, 2002).

While a peer should work in the same field of research as the author and be 'an expert having special knowledge and being a recognized authority' in that field (Burnham, 1990, p. 1325), reviewers of this calibre are not always available. Reviewers may be more experienced than authors in general, but due to the intensified specialization of knowledge, they often know less than the author about the specific work under review. The dilemma is viewed by Henderson (2001, p. 47) as 'the Achilles' heel of peer review'.

1.3.3 The functions of journal peer review

The earliest official board of reviewers was believed to have been established in the 18th century by the Royal Society of London for the 'express purpose' to 'control the quality of the papers accepted' (Wilson, 1978, p. 1697). However, systematic practices of peer review did not figure prominently in early scientific journals because very few scientific papers were submitted (Garfield, 2006). Yet, before long, journals realized the danger of losing public credibility by the indiscriminate publication of all submissions and began to embrace 'quality control' in their practice (Roediger, 1987).

At roughly the same time, scientific knowledge became increasingly specialized and editors began to see the need to seek advice from experts in making decisions about highly specialized papers (Burnham, 1990; Meyer, 2000; Roediger, 1987). As a result,

in the early 20th century, peer review started to take its contemporary form as an organized, institutionalized 'system' in all scientific disciplines (Wilson, 2002).

These three grounds for establishing peer review – quality control, providing expert advice, and establishing credibility – remain the basic functions of peer review for contemporary journals (Bailar & Patterson, 1985; Bingham, 1999; Fox, 1994; Ingelfinger, 1974; Kronick, 1990; Mruck & Mey, 2002) but not without problem.

a) The function of quality control

This function is described by 'sifting and sorting' (Wilson, 2002, p. 164). It is widely accepted that without peer review readers would be faced with a massive number of articles of wildly varied quality, making it hard to detect works of value.

Are some high quality papers wrongly rejected? This question persists and is not easy to answer, due to the difficulties in reaching an agreement on what makes a good paper (Bailar & Patterson, 1985; Smith, 2006a). Furthermore, there are complaints from academics that too many substandard papers have been published (Broad, 1981; Eisenhart, 2002). In addition, in the case of highly innovative works, reviewers may draw on an established paradigm and judge novelty as unsound (Atkinson, 1994; McCutchen, 1991). When this happens, peer review actually takes 'quality control' to an extreme form as 'a screen to limit innovation and enforce conformity' (Wilson, 1978, p. 1698).

Apart from its direct effect of 'sifting and sorting', peer review also contributes to quality control indirectly. Sociologist Robert Merton once claimed that, 'Knowing that their papers will be reviewed, authors take care in preparing them before submission, all the more so, perhaps, for papers sent to high-ranking journals with a reputation for thorough refereeing' (cited in Patterson, 1994, p. 5).

In recent years, the increasing pressure to publish has 'led authors to view publication as the objective of their work rather than the dissemination of knowledge and ideas' (Bence & Oppenheim, 2004, p. 65). Some authors are reported to 'abuse' peer review by intentionally using reviewers to 'finalize' their papers (Fyfe, 1994, p. 62).

b) The function of providing expert advice

There are two issues that may undermine the function of expert advice through peer review. First, manuscripts are not always reviewed by competent peers, so the advice provided may not be reliable (Henderson, 2001). Second, that 'nominated peers are implicitly elevated to assumed superiority in the particular topic of inquiry' may encourage arrogance among reviewers (Atkinson, 1994, p. 150). Reviewers can provide unjustified advice which is then taken into consideration by the editor who may have insufficient expertise (or time) to check it, which in turn undermines the reliability of the editorial decision (Starbuck, 2003). Even when the advice is expert and reliable, authors can be instinctively resistant to criticism, especially when the reviewer's tone is unduly severe. It is known that authors of rejected papers can resubmit their work to other journals without evidence of much revision based on initial reviewers' comments (Tobin, 2002; Wager & Jefferson, 2001; Wilson, 1978).

c) The function of establishing credibility

On the one hand, it is expected that peer review will establish the credibility of 'the science', but there is awareness that this is not always guaranteed. For example, Lock (1994) while editor of the British Medical Journal listed major cases of misconduct in articles, such as reports based on falsified results, published in peer-reviewed journals in the 1980s.

These revelations were followed by other articles that questioned peer review as a reliable means of detecting fraud or misconduct, acknowledging that peer review involves a high degree of trust in the author, and plagiarism and fraudulent behaviours that cannot be detected easily until the work is published and more widely known (Atkinson, 1994; Fox, 1994; McCutchen, 1991; Relman, 1990; Scott, 2007; Smith, 2006b; Wager & Jefferson, 2001; Weller, 1995; Wilson, 2002).

According to Lock (1994), an unrealistic expectation has been built around peer review. Peer review has come to be seen as the 'Holy Grail', while in effect it 'does not and cannot ensure perfection'. Lock emphasises 'the gold standard for the quality of any reported work must remain time' (p. 60).

In addition, there are some functions of peer review that have gained prominence in the last two decades; these include helping authors improve manuscript quality, nurturing collegiality, and defending the autonomy of the academic community.

d) The function of improving manuscript quality

This function was initially contested. In environmental science, for example, Manheim (1975) argued that there was the potential for reviewer comments aimed at improving quality to undermine peer review as the 'arbiter of quality' and cause 'undesirable proliferation of the literature by raising to publishable level manuscripts that could not have been accepted otherwise'; he warned that 'one must carefully distinguish between the unquestioned benefit provided to individual papers and the overall practice in encouraging sloppiness and poor scholarship' (pp. 191-192). Overall this function was noted as not receiving much recognition at that time (Wilson, 1978, p. 1697).

The function began to receive more positive attention by the 1990s (Bingham, 1999; Miser, 1998; Osborne & Brady, 2002; Smith, 2006a; van Rooyen, 2001). From an education perspective, Graue and Grant (2005) note the 'essentially pedagogical' role of peer review (p. 268), while Richard Smith (1999b; 2006a), editor of the British Medical Journal (1992-2004), contended improving manuscripts is one of the essential roles of reviewers.

e) The function of nurturing collegiality

The function of peer review in relation to collegiality and academic culture has also received more attention in recent years. As Chubin and Hackett (1990, p. 84) noted in their book: Peerless science: Peer review and U.S. science policy, peer review nurtures 'communal trust in the publication decision by creating a unique formal consultation among authors, editors, and reviewers'. Osborne and Brady (2002, p. 165, p. 169) viewed this function as the 'true purpose' of peer review in academe:

> [it] is this process of being a reviewer and being reviewed that makes one part of the community… This involves enabling divergent opinions to be heard and this in turn enables the evolution of the field from within, for

the danger is if evolution does not come from within it is imposed from without'.

Knoll (1990, p. 1332) argued that peer review is not 'a truth-grinding machine', but is about 'a discussion among honest and able people, working within the social system of institutionalized science, making the clearest sense they can of the information they all share.'

Peer review is, therefore, a dialogue, but compared to other kinds of academic discussions, it must be seen as an 'attenuated dialogue' with 'the editor insulating authors and reviewers from actually contacting each other' (Bingham, 1999, p. 211). And within this framework there can arise a legitimate process of abusing scientific collegiality. As some editors noted (Rojewski & Domenico, 2004; Smith, 2006a; Wilson, 2002), reviewers may treat authors as potential competitors and block their ideas by rejecting the paper unfairly or responding to the editor late.

f) The function of defending academic autonomy

Advocates for this function argue that peer review acts to protect academic autonomy (often specifically referred to as scientific autonomy). It contributes to the maintenance of the integrity of the knowledge base in a discipline and justifies this to the outside world (Cain, 1999). It protects individual journals from centralized planning and control within a discipline by operating individually and separately (McCartney, 1973a). It is also an 'overwhelmingly 'internalist' process… in which scientific opinions and criteria dominate to the exclusion of others' (Scott, 2007, p. 840). This role of peer review can help explain the deliberate attempt (of editors in particular) to maintain a thread of secrecy in the process. Scott (2007) claims that this is a cost worth paying to maintain peer review as the standard system of selection.

g) The function of stratifying publications in a field

More often than not, a paper rejected by one journal will eventually be accepted by another, often of a lower rank, after several resubmissions (Tobin, 2002; Wager & Jefferson, 2001). This cannot be attributed entirely to the improvement of the paper

because it can be shown that many were not revised before resubmission (Wager & Jefferson, 2001; Wilson, 1978).

What peer review also achieves in the selection process is steering submissions of different quality into different journals, which resembles the act of a 'switch' (Bailar & Patterson, 1985, p. 654) or a 'traffic policeman' (Wager & Jefferson, 2001, p. 261).

In addition to the aforementioned primary functions, peer review is also reported to serve a few less functional (or more symbolic) roles, which resemble democracy and bureaucracy.

A number of scholars have drawn an analogy between peer review and democracy (Finn, 1986; Madden, 2000; McCutchen, 1991; Rennie, 1993; 1999; Robin & Burke, 1987; Smith, 2006a; Spier, 2002). For example, Rennie (1993, p. 2856), while deputy editor of the Journal of the American Medical Association, noted that, 'Peer review is like democracy, which is, to use Churchill's phrase, the worst form of government except all those other forms that have been tried from time to time'. Since the limitations of peer review are so well known to its participants, but no convincing alternative has been found, such an analogy can be taken as a compromise between the capability and fallibility of the process and to maintain its role of upholding the order of scientific communication. Rennie (1999, p. 11) provides an explanation of why a system of peer review that resembles democracy is preferred:

Editors like the comfort of having experts… shoulder the blame for the unpleasant editorial tasks of actually having to make decisions [and] take responsibility for those decisions… authors like the assurance that at any rate some outside experts were called in to moderate the arbitrary decisions… reviewers like the compliment being paid to them in being asked to be included as editors by proxy… and they like having a vote. Finally, readers, who are the majority voters in this democracy, are reassured to find choices made for them as they wade through the information avalanche.

Knoll (1990, p. 1331) noted another, less acknowledged, function of peer review:

> [It is] now a bureaucratic rather than a collegial process… Journals use peer review because it is the way the game is played – not simply because the editors need consultation (although they probably do most of the time). New journals spring up virtually every month, sometimes to cover remarkably tiny subspecialties, and their announcements emphasize that the journals will be peer-reviewed. In cases like this, the use of peer review is more a claim to professional status than an intellectual necessity.

In the context of research assessment and academic benchmarking, peer review has been taken up as 'a marketing tool' for journals to demonstrate 'that their quality control is tight' (Rennie, 1999, p. 10). Although not much discussed in the literature, this function is certainly becoming more common not least in the context of the use of peer review in research assessment exercises and the growing importance of impact scores.

The Australian situation provides a case in point. In 2008, the ARC released a list of discipline-specific, four-tiered, research journal rankings in 22 subject areas as part of the ERA evaluation of research quality (Australian Research Council, 2008). One of the ranking criteria looks at the composition of journals' advisory boards – a journal gains in status if its reviewers are mostly field leaders. So the involvement of advisory board prestige as a ranking criterion serves as a proxy for quality. However, there is evidence that eminent researchers are not always good reviewers (Murphy & Utts, 1994) and prestigious journals do not necessarily offer high-quality reviews (Starbuck, 2005).

The above sections have provided a sense of the issues surrounding journal peer review in connection with its perceived functions, and has identified how these functions have become more diverse as different uses have emerged for, or been imposed on, the activity. Concerns about how well these functions are enacted, particularly the first three, have driven the research agenda in the field for the past fifty years. It is to an overview of this body of work that this thesis will now turn.

1.4 An Introduction to the Empirical Literature on Journal Peer Review

In the initial search for literature on the topic, 385 articles that reported original studies were identified. Table 1.1 presents the disciplinary distribution of the studies. Physics and chemistry and education are presented as separate categories as they are the subject of this study, and it is clear they constitute overall a very small proportion of the total. By contrast, research in medical journals has dominated the research.

Table 1.1 Research Literature Distributed by Discipline Studied[1]

Discipline	Count	%
Business & Economics	32	8.3%
Physics & Chemistry	11	2.9%
Other Science, Information & Mathematics	15	3.9%
Medicine (including Biology)	176	45.7%
Psychology	49	12.7%
Social Science & Sociology	31	8.1%
Education	11	2.9%
Multiple disciplines	60	15.6%
Total	385	100.0%

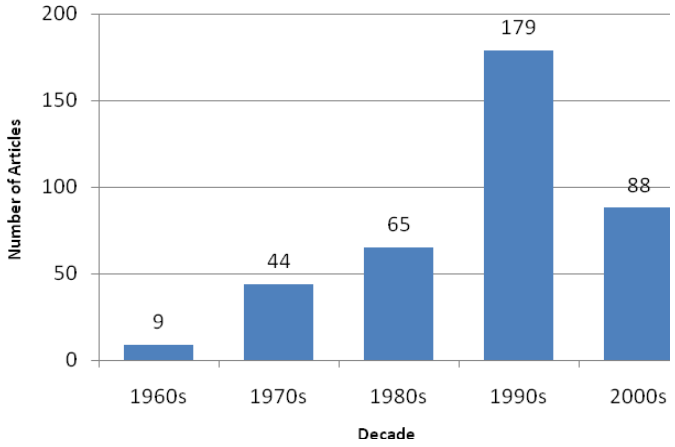


Figure 1.1 Research Literature on Peer Review: Decades of Publication

[1] In many cases, a study was conducted in journals of one discipline, but was reported in a journal of another discipline. This table records the discipline that was the subject of the study.

Figure 1.1 reveals a substantial growth in studies of journal peer review in the 1990s. When the studies were analysed for topic, 61 categories were identified (Appendix 1) which were further grouped into 17 themes (Table 1.2) that were divided into two main areas: 'context and management of peer review' and 'practices and outcomes of peer review'. The first has more of an emphasis toward policy, while the second is more focused on the activity of review. When a separate analysis was undertaken by theme and discipline, the disciplinary pattern was not found to be markedly different from that shown in Table 1.1.

As Table 1.2 shows, the dominant focus of the literature was on the practices of journal peer review (prefixed by the number '2' in the third column). As also indicated in the third column, the review of the literature is organized around the 18 themes in a way to highlight the connection between them on a more or less common sense basis. This is elaborated in Figure 1.2.

The following sections focus on the grouping 'context and management of peer review' (the upper grouping in Figure 1.2). The review starts with a focus on editors and reviewers and their characteristics, roles, and responsibilities. It moves to look at the authorship and the expectations underpinning review and finally at trends and development.

Table 1.2 Main Themes in the Studies about Journal Peer Review

Major Themes	N of Studies	Chapter Section
Reviewer agreement/interrater reliability	66	2.2
Bias in journal peer review	64	2.3
Reviewer comments and recommendations	55	2.1
Recruitment and characteristics of editors and reviewers	50	1.4.1
Blindness and anonymity in reviewing	44	2.4
Reasons for acceptance or rejection of papers	44	2.6
Effectiveness and quality of peer review	42	2.5
Authorship issues	41	1.4.3
Rejection rates	37	2.7

Major Themes	N of Studies	Chapter Section
Efficiency of peer review	33	2.9
Trends in journal publication and peer review	28	1.4.6
Mechanisms to improve peer review	17	2.10
Citation rate and impact factor	17	1.4.8
Fate of rejected papers	15	2.8
Other areas in relation to peer review	15	1.4.7
Rewards of involvement in reviewing and publishing	14	1.4.5
Journal guidelines for reviewers and authors	11	1.4.4
Behaviours of editors and reviewers	9	1.4.2

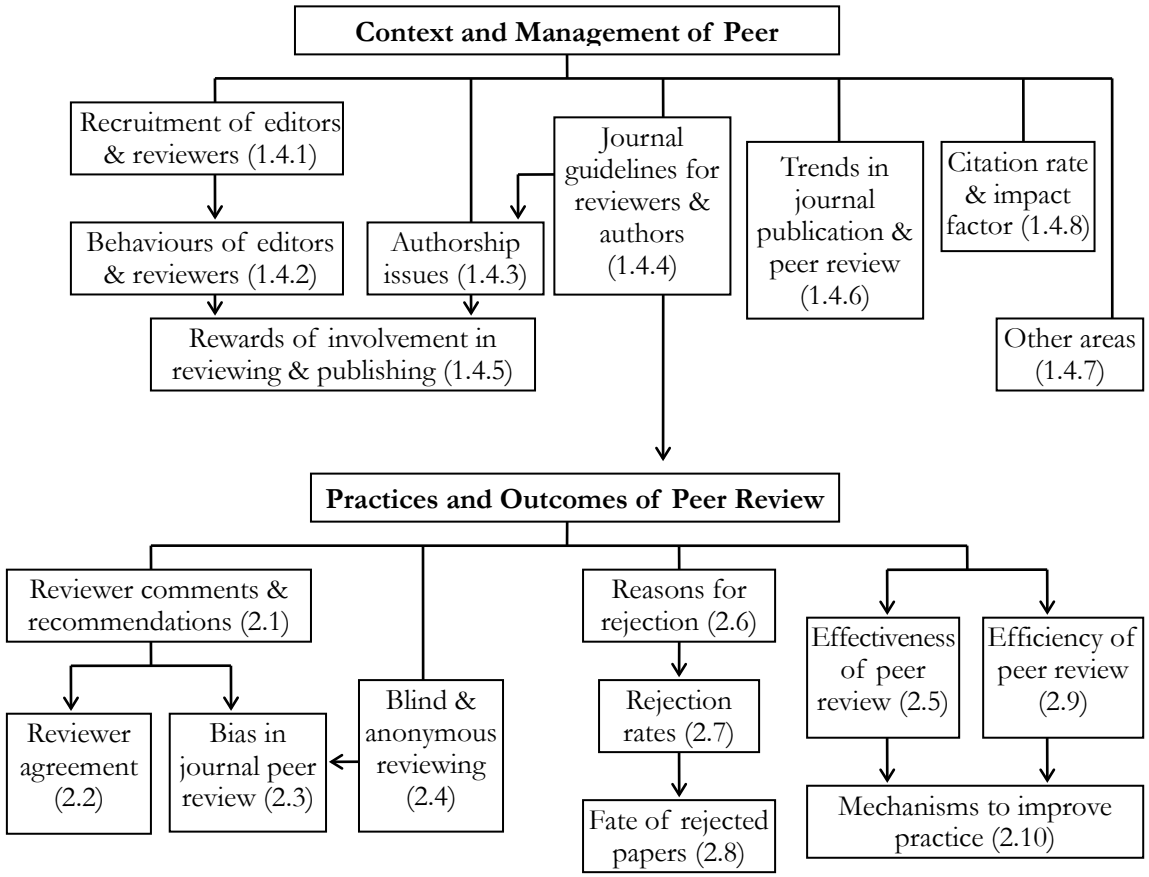


Figure 1.2 Groups and Themes in the Peer Review Literature

1.4.1 Recruitment and Characteristics of Editors and Reviewers

The reviewing process begins with journal editors and is carried out by reviewers. Reviewers are often called the 'gatekeepers of science' for the obvious reason that they take control of paper evaluation and selection (Beyer, 1978; Crane, 1967; Glogoff, 1988; Gordon, 1980; Harcum & Rosen, 1993).

The ultimate responsibility for publication integrity lies with the editor, and so appropriate selection of editors is critical to the validation and construction of the knowledge base of a field (Lindsey, 1992). How are editors selected? According to some early studies, editors typically graduated from prestigious universities (Crane, 1967; Silverman, 1976; Yoels, 1971, 1974), were known for their expertise in a field (Beyer, 1978), and had acted as an editorial board member or reviewer before being appointed as editor of the journal (Silverman, 1976). In a study by Gibbons (1990) no relationship was found between institutional affiliation and editorial board membership.

A more recent study revealed that often editors were approached with a job offer by the retiring editors who they knew personally (Banks & Pracht, 2005).

Several earlier studies in medicine and economics revealed gender difference in that they found a lack of representation of women holding senior editorial posts (Benedek, 1976; Gilbert, et al., 1994; Lock & Smith, 1990). This theme while not common in the literature is raised again in Chapter 4.

In terms of publishing success, editors and editorial members have been found to publish and be cited more often than comparable academics (Braun & Bujdosó, 1983; Lindsey, 1976; Nicholas et al., 2005b; Smith & Dombrowski, 1998; Zsindely & Schubert, 1989; Zsindely et al., 1982). It has been argued that their status and strong networks play an important role in their publication success (Blank, 1991). The quality or impact of editors and board members' publications (measured by citations) is also shown to have a positive effect on the impact factor of their affiliated journals; the extent of which differs by discipline (Braun & Bujdosó, 1983; Lindsey, 1992; Pardeck, 1992).

There are variations in editors' practices of recruiting reviewers. Most editors in Gordon's study (1980) claimed that they were able to select most reviewers purely on the basis of their own knowledge, with or without the input of other editorial staff. Several authors suggested that editors might not necessarily know who the competent reviewers were in a particular area (Bedeian, 2003; Bradley, 1981; Epstein, 1990; Nicholas et al., 2006; Thyer & Myers, 2003). Hamermesh (1994) examined 343 reviewers of seven economic journals and found that editors tended to recruit people they knew, such as colleagues or previous authors of the journal, and researchers whose work was cited widely and who were near the peak of their career.

Gilliand and Cortina (1997) compared the content of 823 papers submitted to a psychology journal with the expertise of their reviewers, and revealed that editors did generally achieve a 'match' between topics and reviewers. In a similar study, Lock (1985) found papers submitted by academic authors were more likely to be sent to academic reviewers and practical based work to reviewers of medical service background.

Editors' criteria for selecting reviewers vary by discipline. A cross-discipline interview-based study found that for physics and biology editors, a reviewer should be up-to-date and competent in the area covered by the paper; for medicine, editors tended to emphasize reviewers' authority in the area; and mathematics editors would use an experienced reviewer to evaluate the overall importance of a paper but draw on a junior reviewer to assess mathematical accuracy (Gordon, 1980).

Some medical researchers explored the practice of asking authors to suggest potential reviewers (Earnshaw et al., 2000; Goldsmith et al., 2005; Gottfredson, 1978; Scharschmidt et al., 1994; Schroter et al., 2006; Wager et al., 2006). They compared author- and editor-suggested reviewers in terms of review quality and recommendation for publication. Although they targeted different journals, their findings consistently showed that the two kinds of reviewers did not differ in review quality, but author-suggested reviewers were less critical and more likely to make favourable recommendations.

Reviewer competence has a lot to do with the quality of what is published. Many researchers have examined reviewer characteristics, to see if they predict review quality. The characteristics comprise educational qualification, experience with research, institutional affiliation, sex, age, and country of residence (Black et al., 1998; Blank, 1991; Braun & Bujdosó, 1983; Callaham & Tercier, 2007a; 2007b; Evans et al., 1993; Gilliand & Cortina, 1997; Godlee et al., 1998; Gordon, 1980; Kliewer et al., 2005; Konradsen & Munk-Jørgensen, 2007; Murphy & Utts, 1994; Nylenna et al., 1994).

The underlying assumption here is that there are factors other than scientific expertise that are considered in the selection of reviewers. However, none of the researchers found any strongly predictive factor in this regard. The only relatively strong factor was age, in that younger reviewers tended to generate higher quality reviews (Blank, 1991; Callaham & Tercier, 2007a; Evans et al., 1993; Godlee et al., 1998; Gordon, 1980; Kliewer et al., 2005).

In an unusual pair of studies (Black et al., 1998; Godlee et al., 1998), two groups of researchers found different results for the same medical journal around the same time. A survey of 670 reviewers of the journal by Black et al. (1998) showed that two factors – being trained in epidemiology or statistics and being external to an editorial board – were positively correlated with review quality, while a study of 420 reviewers on the same variables revealed no significant relationship (Godlee et al., 1998).

Two other studies with reviewers of different medicine journals (Callaham & Tercier, 2007a; Evans et al., 1993) revealed that being younger than 40, on the editorial board, and working in a good university were three significant predictors of reviewing quality, while formal training showed no effect.

In the absence of a definition of a 'good reviewer' and contradictory evidence about good reviewers, it can be difficult for editors to select reviewers or design training for them. This lack of definition, according to Callaham and Tercier (2007b, p. 38), is a 'crucial limitation in the peer review process'.

1.4.2 Behaviours of Editors and Reviewers

The behaviours of editors and reviewers have rarely been a central focus of research in peer review. The available evidence indicates that editors play a core role in the determination of what is published (Braun & Bujdosó, 1983). They do this directly when they decide not to review or not to accept a paper that reviewers have recommended for publication, and indirectly when they select particular reviewers (Crane, 1967). Authors surveyed by Nicholas et al. (2006) criticized editors who acted 'just as administrators'; they noted that editors should not 'leave the ultimate decision of rejection or acceptance of an article to reviewers' (p. 197).

A study by Street et al. (1998, p. 18) was claimed as the 'first systematic, empirical analysis of editor behaviours' at the time. In order to identify positive and negative behaviours, they conducted two focus group interviews with six authors in management and six in business, and then surveyed 111 authors of three top-tier management journals. The survey contained eight positive editor behaviours and 12 negative ones, identified by the focus groups. The authors were asked to rate them for importance and frequency of occurrence; they perceived positive behaviours to occur much more often than negative ones, and that certain behaviours could potentially threaten the validity of peer review, such as editors intentionally selecting reviewers who were likely to review in the direction of the editors' preference. A factor analysis of the survey results also identified a five-factor structure for behaviours – consideration of authors, decision making, thoroughness, ethics, and editorial intervention.

Examination of the behaviours of reviewers appears to be missing in the literature, and seldom have researchers directly explored reviewers' own understanding and execution of their responsibilities. Since reviewers are at the core of journal peer review, this absence of empirical evidence is a major gap in the literature.

1.4.3 Authorship Issues

The opinions of authors are sometimes sought in the literature peer review, but typically studies focus on practices, rather than the experiences. Studies about authorship in the context of peer review range across topics such as co-authorship,

the number of authors per paper, the order in which author names appear on the paper (findings illustrating these trends can be found in Appendices 2, 3, and 4), and honorary and ghost authorship. In other fields, ones with a stronger qualitative research focus such as studies of academic identity and culture, the importance and trials of academic writing and getting published are often explicitly covered, especially in handbooks for new researchers. It needs to be noted here that such work is sparsely if ever referenced in papers on peer review, even though reflection on peer review and its context by editors and others constitutes quite a substantial body of published work.

The growth in co-authorship has been most pronounced in medicine and the sciences, prompted by the growth of collaborative research over the last few decades (Barnett et al., 1988; Borry et al., 2006; Presser, 1980; Terry, 1996), and also growth in the size and specialization of fields in these disciplines (Barnett et al., 1988; Hargens, 1988; Hudson, 1996; McDowell & Melvin, 1983). Laband and Tollison (2000) noted that in some economic journals such growth was more a result of increase in the number of authors per co-authored paper than in the proportion of co-authored papers.

There were three early studies testing the correlation between co-authorship and the probability of publication (Gordon, 1980; Presser, 1980; Zuckerman & Merton, 1971) and revealed a positive relationship between joint work and favourable reviewer recommendation or editorial decision, but interest in this phenomenon may have dissipated because no more recent work could be located. Similarly a number of studies examined citation rates to see if success favoured co-authors over single authors.

Bayer (1982) analysed 110 articles published in a marriage and family journal during 1970-73 and found nearly half of them were multiple-authored and received higher citation rates than those by single authors. Smart and Bayer (1986) further addressed the question in three applied science fields – clinical psychology, management, educational measurement, and found co-authored articles were cited 18 to 64 percent more frequently than single-authored ones.

A study of three economic journals identified a similar pattern (Hamermesh & Oster, 2002), so did a study of 656 articles published in four science journals (Bridgstock, 1991). Overall there was no strong relationship identified between co-authorship and quality.

Several studies have looked at how author order is determined (Goodman, 1994; Hoen et al., 1998; Ilakovac et al., 2007; Shapiro et al., 1994); also see Appendix 4 for more detail. Authors in the studies tended to specify the same top three criteria, namely, authorship was determined by who contributed the 'original idea for the study', 'research design', and 'draft writing'.

The order of authors was not always found to be an accurate measure of intellectual contribution, however, as 'honorary' authorship' proved to be quite common. Supervisors (Sherrell et al., 1989) and those who secured funding for the research were noted as common inclusions (Eastwood et al., 1996). For a variety of reasons heads of departments, laboratories, or research groups who made little contribution to a paper were offered 'gift authorship', (Drenth, 1998; Eastwood et al., 1996; Flanagin et al., 1998; Goodman, 1994; Ilakovac et al., 2007; Marušić et al., 2006; Shapiro et al., 1994; Smith, 1994; Wilcox, 1998). Others were included because of technical expertise (Hoen et al., 1998). Some individuals whose contribution was sufficient for inclusion but excluded were noted as 'ghost authors' (Flanagin et al., 1998; Wilcox, 1998).

Many authors disagree on level of contribution (Hoen et al., 1998; Ilakovac et al, 2007; Marušić et al., 2006; Shapiro et al., 1994). Such evidence highlights the issue of ethical responsibility in academic collaboration that has been taken up vigorously by some journals through formal authorship standards (Wilcox, 1998), although such standards have also generated controversy (Flanagin et al., 1998; Goodman, 1994; Hoen et al., 1998; Ilakovac et al., 2007; Marušić et al., 2006; Shapiro et al., 1994).

Overall, co-authorship remains a largely unstudied and unregulated aspect of journal peer review), creating another layer of potential confusion in the deliberations of reviewers and editors.

1.4.4 Journal Guidelines for Authors and Reviewers

The practice of providing instructions varies widely across journals. Weller (1987) argued that the amount of detail in guidelines for authors was an indicator of journal quality. To test this, she compared the instructions in four categories of medicine journals grouped by prestige, and found a positive relation between the prestige of a journal and the length of its instructions and its likelihood of explaining its peer review process. In contrast, two more recent studies (Altman & Schriger, 2005; Schriger et al., 2005) mapped the content of instruction for authors for 166 and 95 medical journals respectively; both found that the rank of journals was not related, either positively or negatively, with the content of the instructions for authors.

Colaianni (1994) surveyed the editors of 293 medical journals and revealed that not all the journals published instructions for authors in their issues, 28 percent did not publish a statement of peer review, nearly half only 'implied' the use of peer review, and none described variations in the peer review process for different types of papers. Swanson et al. (1991) surveyed the editors of 81 nursing journals and found they published instructions for authors frequently, and some did so in each issue with very specific directions; and there was considerable variation in the length and requirements of the instructions.

Journals also provide guidelines for reviewers, but do they make any difference to the quality of reviewing? A study tested this by comparing the situation of two prestigious medicine journals before and after the implementation of a set of guidelines which contained a checklist of 35 criteria of evaluation (Jefferson et al., 1998). They surveyed 15 editorial staff of the two journals and found that while the editors were aware of the guidelines and believed they were useful, only a minority of the reviewers used the guidelines for their review and no obvious effect was observed.

1.4.5 Rewards of Involvement in Reviewing and Publishing

A question often raised in the peer review debate is why academics participate. It was found that reviewers were typically driven by a concern for the quality of their field

and its scholarly publications (Golde & Walker, 2006; Hamermesh, 1994) and the belief in reciprocal benefit to be obtained from reviewing for high quality journals (Engers & Gans, 1998). Tite and Schroter (2007) revealed that 54 percent of the 551 reviewers they surveyed agreed that financial incentives would encourage them to accept requests to review, but in terms of 'encouragement' to review, 78 percent cited monetary payment alone as the least important factor.

Research also shows that the activity of editing and reviewing is a reward in itself: by accessing and evaluating the work of peers, those involved learn of the latest development in the field, know better how to conduct a rigorous project and write clearly and convincingly. They also develop a connection with other researchers or editors which in turn help them with their own research and publication activities (Braun & Bujdosó, 1983; Campanario, 1996b; Houlihan et al., 1992; Kawano et al., 1993). Reviewers also receive some tangible rewards such as free subscription, annual acknowledgement published in journal issues, and improved opportunity of appointment to the editorial board (Chang & Lai, 2001; Loonen et al., 2005; Tite & Schroter, 2007).

Engers and Gans (1998) argued that just because reviewers would work voluntarily did not rule out the usefulness of payment to inspire speedier and better reviewing. For less successful journals for which many reviewers decline to review, it could be useful to adopt payment to attract more reviewers, and then improve the journal's quality in the long run (Chang & Lai, 2001). Paying reviewers would also speed up the review process (Nicholas et al., 2006).

Authors clearly benefit from peer review. Ziman (1968) noted that journals offered 'speed and permanence of publication for the results of a great many investigations which interact with one another, stimulate further work and form the vast bulk of detailed observations on which major scientific advances are built' (p. 105). Based on the responses of 2500 authors from different fields, Swan (1999) revealed that their main objective for publishing was 'scholarly communication', followed by 'career development', 'personal prestige' and 'funding for future research'. Only a minority cited 'direct financial reward' as a goal. Shaw (1994) surveyed a group of 162

members of two professional associations and while only 32 percent of them had written a paper for publication, when asked about the benefit of publication, they felt it would enhance their professional image and development. Rehr et al. (1998) examined the publication activities of 110 social workers (most being non-academics) and found the top three rewards they hoped to achieve from publication were 'collegial recognition', 'professional prestige', and 'professional knowledge'.

To sum up, it is evident that editors, reviewers and authors all benefit from being involved in journal reviewing and publishing. Peer review is viewed as 'an amazing cooperative network' (McPeek et al., 2009, E155) which contributes to building their academic capacity as well as advancing the knowledge of a discipline.

1.4.6 Trends in Journal Publication and Peer Review

This section discusses a number of studies selected because they show key trends in journal publication and peer review, starting with submission inflation.

Lavelle (1966) surveyed the editors of 166 journals, mainly in the humanities and linguistics and found that during 1961-1966 there was a major growth in the number and circulation of journals (especially those owned by learned societies), and in the number of submissions, in rejection rates and publication lags. The respondents noted that although more papers were submitted, they had not perceived a marked increase in their quality. The inflation of submissions and publication lags was reinforced by many subsequent studies (for example, Bedeian, 2003; Hargens, 1988; Miller & Serzan, 1984; Moyer & Crockett, 1976).

Two consequences of submission inflation are the proliferation of publications and shortage in journal space. Academics responding to a survey by Rowlands and Nicholas (2006) indicated their concern about this and low-quality, scantily peer-reviewed material being published, especially via the Internet. With respect to shortage of journal space, Moyer and Crockett (1976) noted that, from early 1970s, a growing number of finance and economics journals had adopted a policy of charging authors for submission. Beyer (1978) also found this in some journals in physics, chemistry, sociology, and political science.

Guidelines for reviewers were not always in evidence. Juhasz et al. (1975) examined journals from 15 countries and 43 fields and found few provided detailed guidelines. Glogoff (1988) surveyed 110 reviewers of 20 library science journals and found that only half of the reviewers received an evaluation form or specific criteria for evaluation. Miller and Serzan (1984) surveyed the editors of 242 journals listed in the Directory of Publishing Opportunities in Journals and Periodicals and of these only one third of the journals adopted an evaluation criteria form. A later study, by contrast, examined 67 of the 100 top-ranking journals in the ISI in 1989 and found that standard reviewer forms and guidelines were used fairly regularly by these journals, irrespective of discipline (Frank, 1996).

The evolution of electronic reviewing and publication also emerged in the literature. Electronic journals first appeared about 20 years ago, originating from electronic newsletters and electronic conference networks (Weller, 2000). In the 1970s, a series of studies commissioned by the US National Science Foundation estimated the merits of an electronic alternative to print journals (King, 1980; King et al., 1979, 1981; King & Roderer, 1978; all cited in Schauder, 1994, p. 76). These studies predicted that, within 20 years, a majority of articles would be handled electronically, and the main barrier to adopting this was a lack of incentive for change. In biomedical sciences, Wood (1998) found publishers, authors and reviewers enthusiastic for these developments. At present, a huge number of journals have adopted online publishing and reviewing systems, and a range of software tools are available to facilitate this.

Some academics noted that electronic publishing would offer a means for gaining competitive advantage for journals, provide more choices for authors (Schauder, 1994), and improve possibilities for communications for authors and reviewers (Bingham et al., 1998). Others showed concern of 'the perceived lack of prestige of electronic journals' (Gomes & Meadows, 1998, p. 180). Harter (1998) analysed the citations to 39 peer-reviewed electronic journals and found rarely did these have a significant impact in their fields. These studies confirmed the valuable changes that the electronic evolution brought to peer review, but also implied that a product of this evolution, electronic journals, was yet to be a substitute for traditional print

journals, especially due to their lack of impact. As the quality of scholarly journal is typically measured by impact factor, the next section takes a closer look at this topic.

1.4.7 Other Areas in Relation to Peer Review

All topics connected with peer review have not been dealt with exhaustively above, such as work on identification of peer reviewed journals, effectiveness of pre-screening, copy editing, or page charges, although these are far from major themes in the literature. For example, Miller and Serzan (1984) tried to determine specific editorial and reviewing practises that distinguish peer-reviewed journals from non-peer-reviewed ones. They developed 12 criteria for defining peer-reviewed journals and asked 242 editors whose journals were indexed in the Directory of Publishing Opportunities in Journals and Periodicals to report whether their journal followed those criteria. It was found that the mostly practiced five criteria were: using external reviewers (74%), using two or more reviewers (73%), reviewer comments sent to authors (58%), guidelines for authors published regularly (57%), and style guidelines published regularly (50%); in contrast, the other criteria were practiced by only a small proportion of the journals (p. 690).

Later on, by comparing two major index directories, Ulrich's International Periodical Directory and the Serials Directory, Eldredge (1997) noted apparent discrepancies in the identification of peer-reviewed journals in clinical medicine. Eldredge originally hypothesized that the discrepancies could be a result of incomplete coverage in the listings of journals by the two directories, but finally revealed that they were due to the widely varied practices of journals describing themselves as 'peer-reviewed'. Eldredge argued that this reflected a 'widespread confusion about the identities of peer-reviewed journals' among editors (p. 419). Similar results were generated by a replicated study (Bachand & Sawallis, 2003). According to Yvette Diven, director of the product management division of Ulrich, there is a misperception among some publishers that a journal is necessarily peer reviewed if it has an editorial board (cited in Tenopir, 2004, p. 32).

There are also disciplinary differences in the identification of peer-reviewed journals. In education, for example, the adoption of double-blind review was an important

condition for being a peer-reviewed journal (Henson, 1993), while single-blind review remains the norm of many other disciplines.

Authors often experience difficulties when trying to publish. 'Lack of time', 'lack of support and resources', and 'publication lag' were recognized as the major barriers. La Forge and Coelho (1998) found the greatest barrier for some psychology authors was the 'difficulty in responding to reviewers' critiques', followed by 'turnaround time' (pp. 36-37). Swan (1999) added that 'problems in peer review', especially the presence of superficial and unnecessarily hostile reviews, was also a major obstacle. The publication attempt of a cohort of social workers was often hindered by the 'lack of time' and 'on-the-job pressure' (Rehr et al., 1998).

Of the respondents who had not written for publication in Shaw's study (1994), 34.8 percent cited 'no time' as the major reason, 23.3 percent cited 'no encouragement', 22.9 percent cited 'don't know how', while 14.9 percent cited 'no interest' (p. 6). 'Lack of time' was cited by many medical authors as the major reason for not seeking journal publication after presenting an abstract in a conference (Weber et al., 1998).

Another topic is the practice of pre-screening by editors. Editors noted that they would reject a paper if it was not suitable for the journal (Hernon et al., 1993) or obviously 'hopeless' (Stern, 1996). Authors reacted differently to this practice: some thought a strong pre-screening was a way to speed up the process (Mason et al., 1992), while some others felt it unethical for the editor to make a decision without consulting external experts (Von Glinow & Novelli, 1982).

With regard to the effectiveness of pre-screening, Neff and Olden (2006) examined the citation data for articles published in 2000 in 14 biological and ecological journals included the ISI Web of Science. They found that pre-screening significantly reduced the number of unsuitable papers sent to reviewers, and the effect of pre-screening by multiple editorial staff was even stronger.

In another study, Johnson et al. (2007) assigned 263 papers to pre-screening plus external review and another 88 to external review only and compared the turnaround time and reviewers' ratings on the quality of the accepted papers. They found that

while reviewers' ratings were very similar between the two groups, the time to reach a decision was shorter in the case which involved pre-screening, and concluded that pre-screening was an effective strategy for reducing turnaround time and reviewers' workload with minimal impact on the quality of accepted papers.

1.4.8 Citation rates and impact factors

Citation rates and impact factor are not strictly a practice of peer review, but are outcomes of publication, but given the close association between quality and impact needs to be introduced here. Studies that cover the topic typically use them to assess the soundness of published articles, for example by testing the existence of reviewer bias by comparing reviewers' ratings of a paper with the citation rate of the published version.

Citations are the number of times that a publication has been cited in other publications, which may indicate its overall scientific utility (Garfield, 1979). The impact factor of a journal is calculated by dividing the number of citations received in the current year for articles published in the previous two years by the total number of articles published in the journal in those two years (Leimu & Korcheva, 2005). Citation rates are also related to such factors as a field's citation practice, journal circulations, and language and country of publication (Starbuck, 2005).

Citation rate and journal impact factor have been established as the measure of quality in disciplines like medicine, engineering and physics, which usually develop an index of journals and rank them yearly by impact. In some disciplines such as sociology, education, and even areas in medicine, the readiness to accept citation rate and impact factor as the standard measure is low and the position against using it often strongly argued from the evidence. For example, Starbuck (2005) found the citation rates of top-tier sociology journals were not much higher than the bottom-tier ones, which implied that not all articles published in top journals were among the best ones submitted to all journals.

Ten editors-in-chief of seven general medical journals participated in an interview study (Chew et al., 2005) and showed mixed attitudes towards the validity of impact

factor as the measure of clinical impact or the quality of journals and academics. They were mainly concerned about the accuracy of the calculations provided by the Institute for Scientific Information (ISI). They noted the tension between aiming to improve impact factors (for example, by publishing fewer articles) and the threat this posed to maintaining a good relationship with their constituent clinicians.

In addition, 35 authors from a range of disciplines represented in the survey by Nicholas et al. (2006) were quite critical about impact factor, believing that it was not a suitable or fair criterion for evaluating scholarly publications, it discriminated against small disciplines and specialized journals, and that it acted 'as a barrier to the development of science as it led to less research diversity' (p. 199).

Leimu and Koricheva (2005) analysed 228 research articles published between 1975 and 2001 in 53 ecological journals and found that articles received higher citations if they were longer, reporting certain findings, authored by native English speakers, by more than three people, or by people from top-ranking universities. The researchers concluded that, since citation rate was affected by factors other than the scientific merits, its might not be a valid gauge.

Despite the criticisms about the validity of impact, the literature shows that many researchers studying journal peer review have used citation rate and impact factor as a measure of quality when examining the quality of peer review. This is probably due to a dilemma that quality is determined by peer review. When researchers need to test the effectiveness of peer review, they have to find another measure of quality to do so.

A number of researchers used citation as the dependent variable to test the effect of some independent variables on review quality, such as journal rejection rates (Rotton et al., 1993; Yamazaki, 1995), article length (Laband, 1990), co-authorship (Bayer, 1982; Bridgstock, 1991; Smart & Bayer, 1986), use of double blind review (Laband & Piette, 1994a), connection between editors and authors (Laband & Piette, 1994b; Smith & Dombrowski, 1998), and judgments made by reviewers or editors (Opthof et al., 2002a; Opthof et al., 2002b). When exploiting citation rates and impact factors to measure quality for their projects, the researchers have never challenged their validity. This could cast doubt on the validity of those projects and their outcomes.

1.5 Conclusion

Peer review has been in existence for a long time, growing and developing in tandem with the higher education and publishing sectors, serving the role of 'arbiter of quality' and undertaken by many thousands of volunteers. It is deeply embedded in the fabric and social world of academic life – both symbolic and functional. By contrast quality assurance is a recent invention, imposed from outside the academy to monitor activity and facilitate comparison, and for its own operation needs to draw on some of the functions, and use the results, of peer review. Emphasis on quality in research in this latter context appears to have drawn on peer review without adequate recognition of its complexity and of the issues that beset its traditional functions.

With university funding at stake and the pressure to demonstrate that research is of high quality, it is not surprising that there is a growing chorus of concerns and questions about the quality of peer review, how it works and how to ensure it works well. Sigelman and Whicker (1987) claim that, 'virtually everyone in academia seems to have an opinion about how well the peer-review process is working' (p. 495), not least because all have a stake in it.

In the peer review literature, the chief investigators are editors, information and library scientists, and special groups pursuing their own areas of interest. As an outsider to this world, and as a doctoral student with an interest in the quality of education research, what is article quality, and what is in the judgment of this quality, the amount and proliferation of works is daunting. Every study appears to focus on a small element of the whole without much co-ordination, authors dip in and out of the field, and for the most part the impression of the enterprise is that of a juggernaut moving slowly with lots of activity on the periphery.

The researcher came to realize that the volume and layers of activity of the surface represented the many day-to-day challenges for the actors involved, whereas the main principles of operation and faith in the system are in essence its engine. Central to those principles is 'judgment' and this in turn has the capacity to be influenced by

a range of factors within and outside the practice of review, which is why as firm a knowledge as possible of what is known about the operation of peer review became the initial target of the researcher.

There are many categories of investigation in journal peer review but much of the research is fragmentary and sporadic, sometimes contradictory, and in general inadequate to reflect both the extent of the activity and the complexities in both the practices and expectations within and across disciplines.

This chapter has reported the positions taken on peer review by many authors and has reviewed studies that examine the context and management aspects of journal peer review. One of the strongest messages emerging from the literature is there is little agreement on what outcomes should be measured or how this can be done (Bailar & Patterson, 1985; Overbeke & Wager, 1999). This and other challenges set the scene for the researcher's intention to explore the complex phenomenon of peer review from an outsider's perspective.

The next chapter continues the exploration of the literature, focusing on the practices and outcomes of peer review (illustrated by the lower part of Figure 1.2).

CHAPTER 2: REVIEW OF THE LITERATURE ON THE PRACTICES AND OUTCOMES OF JOURNAL PEER REVIEW

2.0 Introduction

In the second part of Chapter 1, the author presented a review of the empirical literature on the context and management of journal peer review. In this chapter, the review focuses on ten themes that emerged from the broader literature regarding the practices and outcomes of peer review (Figure 2.1, chapter section numbers are shown in brackets). It covers the following topics: the outcomes of review (for example, reviewer comments and recommendations, reviewer agreement, reasons for rejection, rejection rates), the process, quality (namely, effectiveness and efficiency) and rigour of reviewing, and potential ways to improve the process.

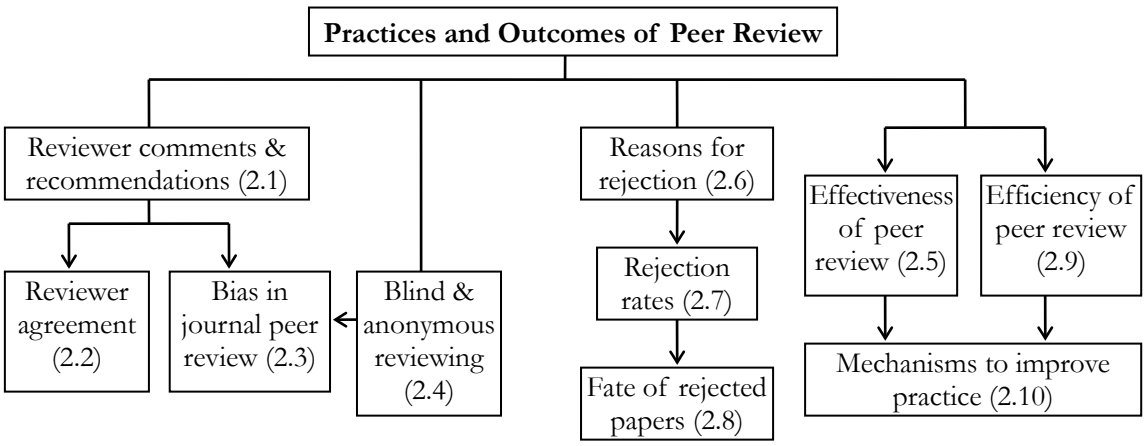


Figure 2.1 Themes in the Literature to Be Covered in Chapter 2

2.1 Reviewer Comments and Recommendations

While there is a substantial body of literature devoted to the practices and outcomes of peer review as will be demonstrated in this chapter, the content of reviewers' reports have attracted relatively little attention compared to areas such as reviewer agreement and bias. The literature reveals that the quality of reviewing is often the target of criticism, especially by authors who believed they were rejected unfairly

(Campanario, 1996a; Simon et al., 1986). A question of considerable importance to all stakeholders is *what makes a good review*?

In order to answer this question, Black et al. (1998) studied 690 reviews prepared on 420 manuscripts for a medical journal. Two editors and the authors were asked to evaluate the reviews against several criteria, namely, the extent to which the reviewers addressed the importance and originality of the research question, strengths and weaknesses of the method, the interpretation of the results and writing presentation, and the extent to which they provided constructive and substantiated comments (p. 232). The researchers did not identify any association between review quality and certain reviewer characteristics and concluded that 'there are almost as many types of good reviewers as there are good reviews' (p.233).

Murphy and Utts (1994) in an analysis of 134 reviews for a biology journal identified no significant difference between reviews prepared by reviewers of different rank, age, experience, or employer regarding the tone, depth, or content of the reviews, with the only exception that reviewers tended to give more constructive comments if their expertise was precisely matched to the topic of the work under review. They also found that the tone of the comments did not necessarily predict the direction of the recommendation.

Studies of reviews have consistently shown the dominance of negative tone (Bakanic et al., 1989; Fagan, 1990; Fiske & Fogg, 1990; Fogg & Fiske, 1993; Runeson & Loosemore, 1999; Tight, 2003) and provide evidence in some areas that the reviews were often contradictory or incorrect (Epstein, 1990, 2004; Peters & Ceci, 1982; Runeson & Loosemore, 1999; Tight, 2003).

In a study of a sociology journal, Bakanic et al. (1989) coded the reviews of 131 accepted and 192 rejected papers into 12 categories. They found that no paper received a completely favourable review; negative comments were much more numerous and more frequently pointed to specific areas than positive comments irrespective of the recommendation. Even so, reviewers were more likely to praise those recommended for publication; and in general positive comments focused more on style, while negative comments more on theory and statistical analysis.

Two studies, one focusing on social work journals and the other on linguistic journals, both revealed that reviewers placed more emphasis on the quality of writing and style than on method, analysis and theory, and were more straightforward and specific in their suggestions for the author if they favoured the papers (Epstein, 1990; Gosden, 2003).

Reviewers demonstrate different styles of reviewing. Cummings et al. (1985) grouped reviewers of a management journal into two styles: coaches and critics, plus those without a clear style. 'Coaches' are those who 'emphasize the developmental aspect of their role' and 'typically reinforce work well-done... promote improvement of subsequent work and provide reasons for recommendations' while 'critics' are those who 'emphasize the evaluative nature of their function' and whose comments are 'almost exclusively critical' (p. 477). It was revealed that the reviewers with a well-defined style displayed more 'good behaviours' than those without this quality (p. 486).

Two similar studies were undertaken later for the same management journal (Beyer et al., 1995; Jauch & Wall, 1989). The one by Beyer et al. (1995) confirmed the findings of Cummings et al. conclusion (1985) about different reviewer styles, and added that the 'coach' style of providing lengthy, constructive reviews was more likely to lead to authors getting published. Jauch and Wall (1989) identified three dimensions in reviewing: 'thoroughness' (being very knowledgeable about the paper), 'substance' (discovering logical or theoretical flaws and giving sufficient reasons), and 'guidance' (providing advice to improve the paper). They also added that regardless of their recommendation most of the reviewers were willing to provide detailed advice for improvement.

In terms of recommendations, reviewers usually make a selection out of four options: 'accept as submitted', 'accept subject to minor revision', 'major revisions', and 'reject'. Two studies found accept-as-submitted was rarely chosen (Linden et al., 1992; Scharschmidt et al., 1994).

In addition, research findings reveal that the final decisions of publication made by editors are closely related to reviewers' comments and recommendation, and the

proportion of favourable comments is far more influential on such decisions than the characteristics of the author or those of the paper (Bakanic et al., 1987; Beyer et al., 1995; Callaham et al., 1998; Fogg & Fiske, 1993; Gilliland & Cortina, 1997; Hargens, 1988; Hargens & Herting, 2006; Linden et al., 1992; Murphy & Utts, 1994; Simon et al., 1986; Smigel & Ross, 1970). These findings reinforce how important reviews are to the final editorial decision.

Some researchers tested the effectiveness of certain journal practices on review quality. Three studies conducted in the same medical journal used a 5-point scale rating on the single item 'review quality' to test the effect of three practices on review quality: editors providing feedback to reviewers on review quality, qualification (namely, university degree, academic status, and experience in research and reviewing), and reviewers attending a 4-hour workshop on peer review (Callaham et al., 2002; Callaham & Tercier, 2007b; Callaham et al., 1998). None of the practices was found to have a clear effect on review quality.

In medicine, an instrument for measuring review quality was developed by van Rooyen et al., (1999) and employed by other studies that targeted medical journals for investigation (Schroter, et al., 2004; van Rooyen et al., 1998b; van Rooyen et al., 1999). By using this instrument, Schroter et al. (2004) tested the effect of training on review quality. The training involved a one-day workshop for one group of reviewers and a self-training package for another. Except for a slight improvement of review quality in the self-taught group, no obvious increase was observed after the training.

The literature across a range of disciplines highlights the variability present in peer review. Reviewers vary widely in their style of reviewing as well as in the focus of their reviews. Moreover, while reviews tend to be dominated by negative comments, they also vary to the extent of being constructive and helpful. What constitutes a good review has proved elusive and there has yet to be any substantive empirical evidence to determine what a good review should look like, let alone a sound measure for review quality. Degree of variation between reviewers on the same paper is also a strong theme in the literature. The next section examines this issue.

2.2 Reviewer Agreement

Reviewer agreement, or 'inter-rater reliability', is typically indicated by the rating or recommendation that reviewers give to the same paper. This topic has almost been exhausted, and gained high profile in the 1990s.

Some studies have tested the overall percentage of agreement, others have looked for the concurrence of general agreement to accept or reject, and some have sought to determine whether the recommendations were within one rating 'step' of each other. Most studies selected reviews prepared for the same journal. A small number compared the results between different journals, for example, Dixon et al., (1983) studied three clinical journals; and Hargens and Herting (1990) studied five journals in two fields, physics and sociology.

Several statistical tests are used to calculate reviewer agreement. Munley et al., (1988) favoured as most appropriate the *intraclass correlation coefficient* which calculates 'the proportion of total variance in review ratings attributed to true differences in papers' (p. 199). A correlation coefficient of less than .40 was regarded as poor reliability (Fleiss, 1981). This test was used by one third of the studies that explored reviewer agreement.

Other tests used include *Kappa* and *weighted Kappa,* which measure overall level of chance-corrected agreement for recommendations (Cicchetti, 1980); *Kendall's coefficient,* a nonparametric test of rank ordering concordance (Feurer et al., 1994); *Finn's r,* which compares 'the variance within papers to the variance that would have been obtained if the ratings had been assigned randomly' (Whitehurst, 1983, p. 74); *single reviewer reliability*, which measures the intraclass correlation of a single reviewer of the same paper across a large number of papers (Marsh & Ball, 1981); *chi square*, which tests for the potential relationship between reviewers' ratings and their characteristics (Murphy & Utts, 1994); and *percent agreement* (Kirk & Franke, 1997), which is the percentage of reviewers who agree on certain items.

One study calculated inter-rater reliability by dividing the number of agreements by the total number of judgments on eight paper characteristics (Kahn et al., 1990).

Some studies also subjected their data to more than one statistical test (Callaham et al., 1998; Daniel, 1993a; Eberley & Warner, 1990; Feurer et al., 1994; Garfunkel et al., 1990b; Hargens & Herting, 1990; Kirk & Franke, 1997; Marsh & Ball, 1981; Morrow et al., 1992; Munley et al., 1988; Plug, 1993; Whitehurst, 1983).

Given the number of statistical measures being used, it is reasonable to assume that inter-rater reliability depends largely on the nature of the instrument. As Kirk (1993) argued, 'How much disagreement there actually is among reviewers depends on how you calculate disagreement and interpret the scores' (p. 3). A good example of this likelihood was provided by Kirk and Franke (1997). They tested the level of reviewer agreement for a social work journal by using four measures: percent agreement, one-step agreement, intraclass correlation, and 95% confidence intervals (p. 122), and found while all four measures suggested low agreement, the result of one-step agreement was double that of percent agreement.

The statistical tests used and outcomes generated by the studies of reviewer agreement are tabulated in Appendix 5. It clearly shows that reviewer agreement was low in most cases, irrespective of discipline. The only exceptions were two studies in medicine (Feurer et al., 1994; Oxman et al., 1991) where high intraclass correlations of 0.84 and 0.71 were revealed, which indicated satisfactory agreement between reviewers. However, the results were not easily generalisable beyond the context due to a limited research design. There were also two studies which revealed higher than 80 percent reviewer agreement for overall recommendations (Boice et al., 1984; Garfunkel et al., 1990b), but their sample sizes were small (25 and 40 papers respectively).

The table of published findings presented in Appendix 5 also shows that the degree of reviewer agreement was higher in physics, chemistry, medicine, and psychology than in humanities and the social sciences, but few attempts have been made to examine such disciplinary variations. Available cross-disciplinary studies emphasized similarities rather than differences across disciplines when interpreting findings (for example, Beyer, 1978; Cicchetti, 1991; Hargens & Herting, 1990; Gottfredson, 1978; Starbuck, 2005; Whitehurst, 1984).

A number of studies revealed that even if reviewers made the same recommendation, they could be basing their judgment on very different grounds (Daniel, 1993b; Fiske & Fogg, 1990; Harcum & Rosen, 1993; Justice et al., 1994; Linden et al., 1992; Lindsey, 1999; Marsh & Ball, 1981; 1989; Munley et al., 1988; Peters & Ceci, 1982; Runeson & Loosemore, 1999; Scharschmidt et al., 1994; Smigel & Ross, 1970).

Several factors were associated closely with reviewer agreement. For example, reviewers tended to agree more on recommending rejection than on recommending acceptance (Cicchetti, 1991; Daniel, 1993a; 1993b; Eberley & Warner, 1990; Hargens & Herting, 1990; Howard & Wilkinson, 1999; Miller & Perrrucci, 2001; Warner et al., 1985), and on judgment regarding the contribution of the research (Cicchetti & Eron, 1979; Dixon et al., 1983; Gilliland & Cortina, 1997; Gottfredson, 1978; Kirk & Franke, 1997; Munley, et al., 1988; Oxman et al., 1991).

Low-level reviewer agreement has been cited as evidence that peer review does not work properly (Ernst et al., 1992; Lindsey, 1988; 1991). This position has generated interest in ways to improve inter-rater reliability, such as by defining more specifically the features of a good review, training editors, and using more reviewers per paper (Callaham et al., 1998; Lindsey, 1988).

Many have cautioned that inter-rater reliability should not be confused with the reliability of peer review, and argued that to improve reviewer agreement should not be a primary goal of editors (Daniel, 1993b; Gilliland & Cortina, 1997; Marsh & Ball, 1989; Scott, 1974).

According to Gilliland and Cortina (1997), low inter-rater reliability could mean that when considering a paper, reviewers focused on different aspects, drew on different interpretation of the same criteria, or demonstrated favouritism toward or against certain characteristics of the author or the paper. As previously discussed, editors may have deliberately selected reviewers of different expertise who would focus on different aspects of the paper in their reviews (Eberley & Warner, 1990; Fiske & Fogg, 1990; Gilliland & Cortina, 1997; Kirk & Franke, 1997; Morrow et al., 1992; Munley et al., 1988; Scott, 1974). Diversity in the reviews thus can actually be desirable. As one veteran editor claimed, 'One of an editor's best safeguards against

routinization and its partners, bias and diminution (one's own and one's reviewers), was low agreement between reviewers' (Glidewell, 1988, p. 766).

Most of the studies of reviewer agreement have based their examination on reviewers' recommendations and ratings of manuscripts. There is another type of reviewer agreement that is little explored, namely, the extent to which reviewers agree on the criteria they use for evaluation and how consistently they apply them in making a judgment, namely, whether and how reviewer perceptions are 'different notions on different things with the same label' (Harvey & Green, 1993, p. 10). It can be argued that this may better explain the differences between reviews without abandoning it altogether a need for some level of agreement. This is one of the key issues that this study has attempted to address.

2.3 Bias in Journal Peer Review

Reviewers bring a mix of preferences, skills, moods, and intellectual 'vagaries' to the reviewing task (King et al. 1997). Hence given the high stakes nature of peer review and how the acceptance of articles defines the body of literature (Cummings et. al., 1985), concerns about ethical conduct of reviewers and reviewer subjectivities have attracted direct attention in a large body of literature. The main topics of reviewer bias in these studies fall into three groups: characteristics of authors; characteristics of papers and bias in the editorial decision-making process. Another group of studies is devoted entirely to the influence of 'blindness in review' (see next section).

Bailer and Patterson (1985, p. 655) defined reviewer bias as 'the use of criteria other than strict scientific and technical merit in framing comments and advice to the editor'. One much-discussed source of bias is related to whether the theory and ideology reflected in a paper conform to those held by the reviewers. Gordon (1980, p. 271) referred to this as 'deliberate bias'.

Several studies tested the potential influence of the latter, all by using bogus or fabricated papers as stimulus, and revealed that reviewers tended to favour work that was congruent with their own theoretical or philosophical views (Abramowitz et al., 1975; Ernst et al., 1992; Mahoney, 1976; Peters & Ceci, 1982; Weller, 2001).

A raft of studies across disciplines has explored bias as it relates to the presentation of statistically significant results in a paper, namely, 'statistical bias'. A number of studies found that reviewers were much more likely to reject a paper if it tested statistical significance but failed to produce statistically significant results, and vice versa (Atkinson et al., 1982; Beyer et al., 1995; Coursol & Wagner, 1986; DeVaney, 2001; Dickersin et al., 1987; Easterbrook et al., 1991; Epstein, 1990; 2004; Fagley & McKinney, 1983; Kerr et al., 1977; Moher et al., 1994; Rowney & Zenisek, 1980; Scherer et al., 1994; Stern & Simes, 1997).

Papers reporting statistically significant results were also published more quickly than those with non-significant results (Misakian & Bero, 1998; Stern & Simes, 1997). A survey of 609 psychology academics (Coursol & Wagner, 1986) and another survey of 156 medicine authors (Dickersin et al., 1987) both revealed that researchers were hesitant to submit papers that reported non-significant results.

Only three studies on the topic were located that did not detect the evidence of this bias (Bero et al., 1994; Lee et al., 2006; Olson et al., 2002). One (Bero et al., 1994) examined journal articles that tested the health effect of tobacco exposure and found that 80 percent of them reported positive results (namely, tobacco smoke was harmful); and of those using statistical tests 57 percent reported statistically significant results and 43 percent did not, suggesting no statistical bias. However, Bero et al. (1994) attributed this to the public preference for negative results (namely, tobacco smoke was not harmful), indicating the role of specific cultures of reporting.

The broader impact of research method on review is revealed in a number of articles. For example, in the 1990s when qualitative research was becoming more common in traditionally non-qualitative fields, it was found reviewers rejected significantly more papers reporting qualitative studies than those reporting quantitative studies (Bakanic et al., 1987; Blank, 1991; Kerr et al., 1977; Patterson, 1994; Staw, 1985; Wallace & Fleet, 1998). Rarely was the opposite reported (Lee et al., 2006). In addition, prospective trials were found to be more likely to be accepted than retrospective studies (Autorino et al., 2007; Kerr et al., 1977), and papers involving sophisticated statistical techniques or extensive documentation were preferred over those without

(Bayer, 1982; Beyer et al., 1995; DeVaney, 2001). Medicine reviewers in particular tended to favour papers reporting randomized controlled trials (Dickersin, et al., 1987; Lee et al., 2006; Scherer et al., 1994).

Some studies explored reviewers' attitudes toward new ideas, unconventional methods, and unconfirmable results. Most of these revealed a stronger predisposition for conventional design or findings over unorthodox ones (Armstrong, 1996; Calnan et al., 2006; Epstein, 1990; 2004; Fölster, 1995; Gans & Shepherd, 1994; Ginsberg, 1999; Mahoney, 1977; Resch et al., 2000; Runeson & Loosemore, 1999). However, the successful test of a new theory or the use of new methods or material has also been reported among the major reasons for acceptance (Hernon et al., 1993; Kerr et al., 1977; Rowney & Zenisek, 1980).

While novelty is one of the key reasons identified for acceptance, the lack of it does not automatically lead to rejection. One study indicated that reviewers indicated their preferences in another way, namely, by putting effort into feedback. Beyer et al. (1995) analysed a group of papers submitted to a management journal, the reviews and the editor's letters to the authors. The study identified only moderate support for the idea that editors and reviewers favoured papers that claimed novelty, rather, the best predictor for acceptance was the extent to which reviewers and editors tried to help authors improve their work.

Another factor that impacts on reviewers is shown to be the reporting of replication of previous studies. Reviewers were found to be reluctant to accept replication studies (DeVaney, 2001; Kerr et al., 1977; Robinson & Levin, 1997). An argument against replication was that journal space was so limited that only original research could be accommodated (Armstrong, 1996). Papers which had been presented previously at conferences and reproduced in proceedings were also likely to receive negative reviews for not being 'original' (Kerr et al., 1977; Rowney & Zenisek, 1980).

Another factor impacting on bias is knowledge of the authors (including status, research experience, gender, employing institution, and country of residence). It has been known from an early period that authors from higher-ranked universities are

more likely to receive favourable reviewer comments (Bakanic et al., 1987; Crane, 1967; Gordon, 1980; Hernon, et al., 1993; Peters & Ceci, 1982).

The likelihood of acceptance has also been found to be associated with authors' country of employment insofar as papers written by those from western, developed, English-speaking countries have been revealed to be more easily accepted (Autorino et al., 2007; Hernon, et al., 1993; Lee et al., 2006; Link, 1998; Loonen et al., 2005; Nylenna et al., 1994; Olson et al., 2002; Opthof et al., 2002a; Ross et al., 2006). Authors' experience as editorial board members or reviewers for a journal may also increase the likelihood of their work being accepted by that journal (Bakanic et al., 1987; Medoff, 2003).

With regard to authors' gender, two studies examined papers submitted to a medicine journal and a psychology journal respectively; in both settings author identity was known to reviewers (Gilbert et al., 1994; Gilliland & Cortina, 1997); neither of these revealed marked reviewer bias on the basis of authors' sex. Another two studies of economic journals employing double-blind review revealed that, under this system, there was no bias against female economists, and the acceptance rates for female authors were slightly higher than those for males (Ferber & Teiman, 1980; Laband, 1987).

A study of 67,275 abstracts accepted or rejected by a medical conference found no evidence of bias against sex in either open or blind reviewing (Ross et al., 2006). Lloyd (1990), however, had detected same-sex preference in five behavioural science journals, and that female reviewers accepted significantly more female authored papers and male reviewers accepted more authored by males.

The seniority of authors in terms of age and academic status (for example, professors compared to early career academics) was not identified to have any strong, consistent effect on successful publication (Bakanic et al., 1987; Beyer et al., 1995; Daniel, 1993b; Miller & Perrucci, 2001; Zuckerman & Merton, 1971).

The foregoing studies have focused on biases in evidence at the reviewing stage. Is bias in evidence in editors? As indicated in Chapter 1, editors do express preferences

in their editorials. DeVaney (2001) surveyed 194 editorial staff of 24 journals in education, sociology, and psychology. The majority of the respondents noted that their journal did not have a written policy about the reporting of results of statistical significance tests and reported they were willing to accept statistically non-significant results. However, more than half of them also believed that papers involving such tests 'should contain statistically significant results' (DeVaney, 2001, p. 316).

With respect to replication studies, the attitude of editors was found to be affected by the field from which they came. In DeVaney's study (2001), most of the responding editors showed willingness to accept replication studies, while some social and behavioural science editors noted that replications were not 'examples of research encouraged for submission in the editorial policy' (Neuliep & Crandell, 1990, p. 87). In a report on two surveys by Madden et al. (1995), natural science editors were found to generally endorse replications as a necessary and even a significant part of research, while social science and marketing editors perceived replications as indicators of lack of originality and seldom published such work.

Some researchers have explored the possibility that relationships between editors and authors could affect the editors' disposition. While an early study in psychology (Schaeffer, 1970) revealed that a majority of the articles under study were written by authors with no immediate affiliation with the editors others have found the opposite. Laband's (1985) study showed that published articles by economists from the same school as the editor were much longer than those authored by unaffiliated economists; and editors tended to use their academic network to capture high quality papers, where articles with an author-editor connection were cited more than those without the connection (Laband & Piette, 1994c; Rodman & Mancini, 1977; Smith & Dombrowski, 1998; Smith & Laband, 1995). On a different tangent, some editors have demonstrated 'criticism bias' in that they tended to be 'looking for a reason to reject' from reviewers (Bedeian, 2003, p. 336).

To summarise, bias in peer review takes two main forms: paper-related bias and author-related bias (or 'content particularism' and 'social particularism', according to Gilliland & Cortina, 1997). Some attributed the former to the absence of precisely

defined criteria of what identifies a high-quality paper in a field (Fölster, 1995; Mackie, 1998), and others have suggested content particularism is more likely to operate in journals from fields where research paradigms are less neatly defined (Epstein, 1990; Pfeffer et al., 1977). As for social particularism, it is widely recognised that knowing the identity of authors is most likely to lead to bias. In the hope of diminishing this type of bias, many journals have adopted 'double-blind review' and it is to the issue of blinding to reduce bias that the chapter now proceeds.

2.4 Blindness and Anonymity in Reviewing

In a single-blind review, the identity of the author is not revealed to the reviewer; in an anonymous review, the identity of the reviewer is not revealed to the author; and in a double-blind review, the author and the reviewer are blinded to each other. In some studies, the practice of non-anonymous review, where reviewers sign their reviews, is also referred as 'open peer review'. The practice of blind and anonymous review was another aspect of peer review that was also exhaustively studied.

Bachand and Sawallis (2003) looked at 533 journals in 18 fields and revealed that 58 percent of the journals used double blind review, 37 percent used anonymous review, and 5 percent released identity mutually; as far as discipline is concerned, double blind was used by all the social work journals and more than 90 percent of the history and business journals, while single blind was used by 90 percent of the biology and chemistry journals and more than half of the medicine and psychology journals.

Freda and Kearney (2005) identified that 95 percent of then existing nursing journals employed double blind review. In earlier studies, Miller and Serzan (1984) surveyed 242 journals of 12 fields and found double blind review was more likely to be used by journals in education, sociology, history, psychology, and philosophy than journals in natural and physical sciences, whereas Budd (1988) in library science and Cleary and Alexander (1988) and Weller (1991) in medicine found only a small proportion of the journals being studied used double blind review.

Some editors in the aforementioned study by Weller argued that there was value in reviewers knowing the authors' identity and some noted that since it was impossible to hide the identity of all authors 'in order to treat everyone fairly', all authors' names should remain on the papers (Weller, 1991, p. 99). The argument raises two questions: what is the value of blind and anonymous review, and how successfully can author identities be concealed from reviewers?

In terms of the effect of blinding identity (of either reviewers or authors), there is both supporting and opposing evidence. By comparing blind/anonymous reviews with non blind/anonymous reviews, little impact of blinding was detected on the quality of reviews, the severity of reviewers' recommendations, time taken to review, the number of errors detected by reviewers, or the reasons cited for rejection (Coe & Weinstock, 1967; Crane, 1967; Godlee et al., 1998; Justice et al., 1998; Parker, 1986; van Rooyen et al., 1998a; van Rooyen et al., 1998b; van Rooyen et al., 1999).

Others studies found that in double blind review, reviews were of significantly higher quality, took longer to complete, and were less abusive than anonymous reviews; in addition, female authors tended to receive higher acceptance rates; and reviewers were less likely to recommend rejection (Bingham et al., 1998; Edwards & Ferber, 1986; Evans, et al., 1993; Ferber & Teiman, 1980; Fisher et al., 1994; Fölster, 1995; McNutt et al., 1990; Parker, 1986; Polak, 1995; Posey, 2005; Ross et al., 2006; Walsh et al., 2000).

One study also showed that, once published, articles receiving double blind review received significantly more citations than those receiving single blind review, suggesting that double blind review helps identify papers of authentically high quality (Laband & Piette, 1994a).

Double blind review is not always truly 'blind' and reviewers can guess the identity of authors (Juhasz et al., 1975; Kearney & Freda, 2005; Weller, 2001). Appendix 6 presents the percentage of correct guesses made by reviewers across studies. At least 25 percent of the authors were identified; in one study it was as high as 68 percent (Justice et al., 1998). Authors were identifiable by reviewers' awareness of their work, for example, an earlier version was presented at a conference, or by authors' frequent

self-citations (Ceci & Peters, 1984; Cho et al., 1998; Fisher et al., 1994; Katz et al., 2002; McNutt et al., 1990; Yankauer, 1991). Easily identifiable authors were often those who were 'better known or who belong to networks that distribute their working papers more widely' (Blank, 1991, p. 1052).

One study turned the framework around and asked authors who had submitted to a specialized medical journal to guess their reviewers (Wessely, 1996). Only 5.9 percent of the reviewers were correctly identified, usually by authors whose paper was accepted. Although the only attempt of its kind, this study revealed that it was probably easier to hide the identity of reviewers than that of authors.

The attitude towards blind or open peer review was mixed among editors, reviewers, and authors. Of the 871 medicine academics who responded to Gidez's (1991) inquiries 'Should reviewers be anonymous?' and 'Should names of author and institutions be withheld from reviewers?' (p. 84), 73.2 percent felt reviewers should be anonymous while only 38.9 percent felt that author should be anonymous.

In a survey involving 88 nursing journals (Kearney & Freda, 2005), 80 percent of the editors viewed double blind review as important; when asked whether they could see any benefit of unblinding author or reviewer, 53 percent said no, 24 percent said yes, and the rest were unsure. Most of them were against open peer review, with some claiming that 'reviews would be less thorough or rigorous if reviewers were concerned about being identified' (p. 448). A survey of 370 chemistry editors and editorial board members also revealed that there were different attitudes among the editors and members regarding going totally blind or totally open (Brown, 2007).

For reviewers, studies have shown that anonymous review was clearly the preferred model (Bingham et al., 1998; Fabiato, 1994; Posey, 2005; Yankauer, 1991), and reviewers who were asked to sign their reviews were significantly more likely to decline to review (van Rooyen et al., 1999a). Most of the reviewers in the study by Fisher et al. (1994) indicated that neither the quality nor the difficulty of their review would be affected by being blinded to the author. In contrast, authors were more in favour of open peer review. They felt disadvantaged in the anonymous system and

believed that an open system would motivate reviewers to take more responsibility as well as facilitate communication between them and the reviewers (Bingham et al., 1998; Fontaine, 1995; Nicholas et al., 2006; Posey, 2005).

To put it simply, reviewers prefer anonymity and authors prefer openness, both for obvious reasons. However, editors do not show a clear-cut attitude. If a continuum of editorial practices is formed from being 'totally open' to 'totally blind', then editors are located along the continuum. As the literature does not provide clear evidence regarding the relationship between different editorial practices and review quality, the choice between anonymity and openness would appear to be hard to justify.

2.5 Effectiveness of Peer Review

Peer review is subject to a range of problems, from the quality of reviewers to the quality of their reviews, and no 'best' solution has been found so far to tackle them. A number of studies have focused on the topic of the 'quality' of peer review. It needs to be noted that studies about reviewer agreement and bias (presented above) need to also be seen as integral to this topic.

Peer review was initially introduced into journals for the purpose of quality control (Burnham, 1990). A few studies have explored the effectiveness of this function by looking at the 'fate' of originally rejected papers that were published elsewhere later (Abby et al., 1994; Daniel, 1993a; 1993b). Abby et al. (1994) revealed that, of 125 papers rejected by a leading surgical journal, 77 were not published by an indexed journal in the three years after initial rejection; of the 48 finally published papers, most were accepted by journals with a lower circulation or citation rate than the initial submitting journal.

Another two studies based on the same chemistry journal found that, not only were most of the rejected and resubmitted papers published in journals of much lower rank or circulation, but they were minimally cited after publication (Daniel, 1993a; 1993b). Based on the findings, the researcher concluded that peer review was effective as a screening process. Supporting evidence was also found in Dixon et al. (1983), Ghali et al. (2002), Loonen et al. (2005), and Opthof et al. (2002a).

In contrast, there were researchers who challenged the reliability of peer review in selecting appropriate submissions for publication. Reviewers' ability to identify flaws in a paper was usually used as an indicator of reliability (Callaham et al., 1998). One study identified errors in accepted papers (Garfunkel et al., 1990b); some others detected errors in published papers, regarding research conduct, statistical analysis, data interpretation, tables, figures, quotation, and referencing (Eichorn & Yankauer, 1987; Evans et al., 1990; Godlee et al., 1998; Gore et al., 1977; Huxley, 1986; Runeson & Loosemore, 1999; Schriger et al., 2006).

Peters and Ceci (1982) selected 12 articles published in 12 leading psychology journals, changed the authors' names and institutions, and sent them to reviewers as if they were new submissions. Only three articles were recognised as being previously published, eight were rejected with strong criticisms, and only one was accepted. Similar findings were produced by a study in management (Runeson & Loosemore, 1999). In another study (Schroter et al., 2004), 609 reviewers were sent a fabricated paper in which nine methodological errors were added deliberately, but only 2.6 errors were identified by the reviewers on average. Campanario (1995, 1996a) from library science and Gans and Shepherd (1994) and Starbuck (2005) from economics and management also found that some highly-cited articles had been rejected initially; the researchers cited this as the evidence against the effectiveness of peer review.

Journal peer review is believed by many academics to help improve the quality of submissions (Beyer et al., 1995). Several studies tested this function by comparing initial submissions with their published versions. The studies all provided supporting evidence that peer review would help improve the presentation and readability (Biddle & Aker, 1996; Jean-Pierre et al., 1996; Metoyer-Duran, 1993; Roberts et al., 1994), the reporting of statistics (Cobo et al, 2007), the accuracy of content (Purcell et al., 1998), and the overall quality of a paper (Campanario, 1996a; Goodman et al., 1994; Smigel & Ross, 1970).

In regard to reviewers' willingness to assist authors, Gosden (2003) analyzed 40 reviews of 21 papers submitted to a physics journal and found that all comments, even those highly critical in tone, showed reviewers' commitment to helping authors.

A number of studies have sought the author's judgments on the effectiveness of peer review of their work. Bedeian (2003) surveyed 173 authors of two management journals and revealed that 89 percent of them believed the 'net effect of peer review' (p. 333) was to improve the quality of the submissions, 92 percent felt editors and reviewers had read their work carefully, and 72 percent felt the critical comments they received were useful. A group of social work authors also rated the helpfulness of reviewers' comments as high (4 out of 5) (Thyer & Myers, 2003).

Another relatively recent and atypically large cross-disciplinary survey of 5,513 authors found that the authors' experience with peer review was generally highly positive, with 77 percent agreeing that the reviews on their last published paper were useful (Rowlands & Nicholas, 2006). Notably, in this study, authors from physics, a field described as 'a community that works within a very open, collaborative, information culture' (p. 39), were the most critical about their experience with peer review.

In these studies, although most of the respondents showed a generally favourable response to, and faith in, peer review, the voice of unsuccessful authors was underrepresented. It was well-known that authors of accepted papers tend to perceive peer review as more beneficial than those authors who are rejected (Bedeian, 2003; Bradley, 1981; Cofer, 1985; Garfunkel et al., 1990a; Rowlands & Nicholas, 2006; Sweitzer & Cullen, 1994; Weber et al., 2002; Weller, 1996). It was also evident that many authors felt pressure when they had to make changes to their work to conform to editors' or reviewers' personal preference or make changes that they felt were wrong, and there were authors who believed that some reviewers contrived their reviews to impress the editor rather than identify the real problem of the paper (Bedeian, 2003; Bradley, 1981).

In summary, the literature presents evidence for the effectiveness of two core functions of peer review: controlling and improving quality for publication. Yet there was also evidence of deficiencies in performance on this score, identified typically as the consequence of poor performance by certain reviewers. Moreover, even with good advice some authors could still struggle to get published (see Section 1.4.7), which implies that peer review's contribution to improving quality is always relative.

2.6 Reasons for Acceptance or Rejection of Papers

Conforming to reviewers' requirements for revision has been reported as frustrating by authors (La Forge & Coelho, 1998), so what of rejection? Available empirical evidence on this topic was obtained in a number of ways, including by analysing reviewers' reports or editors' decision letters, asking editors or reviewers to reflect the criteria they used in making such decisions, or ranking the importance of a given set of criteria by means of survey or interview. Compared to the other themes reported above, literature on this topic is spread more evenly across disciplines.

In the early 1990s, Campion (1993) invited 191 academics in the natural and social sciences to nominate criteria they used to evaluate academic writing. From this list he identified 15 categories and 43 sub-categories that together provided a detailed account of what a high quality paper should look like. By contrast, reviewers were shown to draw on only a few criteria to reject a paper or recommend major revision (Bedeian, 2003; Jauch & Wall, 1989; Marsh & Ball, 1989; Runeson & Loosemore, 1999).

How reviewers weight the criteria they use or what they prioritize remained under-researched, even though a number of studies have collected information on the experience and perception of a general group of academics (Jauch & Gentry, 1976; Sternberg & Gordeeva, 1996; Sutherland et al., 1993).

Gordon (1980, p. 263) argued that 'editors differ in their view of the role of journal publication within the social and intellectual activity of their discipline, and so in the criteria they seek to satisfy as preconditions to the acceptance of papers'. Beyer (1978) compared the rankings of criteria by editors of four fields and found editors in the physical sciences ranked 'contribution to knowledge' most highly, and editors in sociology and political science ranked 'logical rigor' as the top criterion for evaluation. Some agriculture editors rated 'contribution to knowledge', 'topic selection', and 'creativity' as the top criteria (Lacy & Busch, 1982).

A group of 79 editorial board members of six personnel guidance journals rated 'contribution to knowledge', 'research design', and 'objectivity in reporting' as the top three criteria (Frantz, 1968). Another study with 132 clinical psychology editors

cited the same three top criteria and added that very high quality papers were differentiated from average ones by 'topic selection', 'theory', and 'style' (Wolff, 1970).

While editors' perceptions about the core criteria for evaluation were consistent, more specific and more distinct features were revealed as reason for rejection. In an earlier study, a group of economics and business editors rated 'no significant addition to knowledge' and 'too superficial' as the most common reasons for rejection (Coe & Weinstock, 1967; 1968). Nursing editors in Swanson et al.'s study (1991) chose 'poorly written' as the most popular reason for rejection, followed by 'poorly developed data' and 'not consistent with purpose'. The top three reasons for outright rejection cited by some higher education editors were 'author guidelines not followed', 'not thorough', and 'bad writing' (Noble, 1989).

Editors would also reject a paper commended by all the reviewers, if its topic had been discussed to the point of saturation in the journal, but they would accept one that was likely to spark debate or controversy (Hernon et al., 1993; Weller, 1991). Some studies analysed editors' letters to authors. The top reasons for rejection cited by the editors of a surgery journal were 'weak conclusion and discussion' and 'poor presentation' (Abby et al., 1994), and for some sociology editors the top reason was 'inappropriate for the journal' (Willis & Bobys, 1983).

Compared to editors, reviewers appear to draw on slightly different reasons for recommending acceptance or rejection. The reviewers of 19 leading social science and management journals in the 1970s rated 'strong author reputation', 'successful test of the author's own new theory', and 'content different from that traditionally published by the journal' as the factors being mostly associated with acceptance, and 'statistically insignificant results' and 'replication studies' as the most popular reasons for rejection (Kerr et al., 1977, p. 141).

Kerr et al.'s instrument (a rating scale consisting of 37 criteria) was adapted later by two similarly designed studies in psychology (Coelho & La Forge, 2000; Rowney & Zenisek, 1980). Both studies confirmed that 'representing a new, original theory' was the major positive factor. Rowney and Zenisek found the most influential negative factor was 'experimental data with no control group' and, according to Coelho and

La Forge, it was 'direct replication'. Reviewers of 31 library journals ranked 'validity of claims' and 'originality' as the top positive factors (Glogoff, 1988).

Some studies also examined evaluation criteria by analysing reviews and extracting the actual reasons that reviewers used to justify their judgment. Most of the reasons for rejection (see the table in Appendix 7) fell into the four areas specified by Kassirer and Campion (1994), namely, research design, writing, interpretation of data, and the significance or contribution of the research.

Some studies provided an interesting comparison between the reasons for acceptance and rejection for the same group of journals (Bakanic et al., 1989; Hernon, et al., 1993; Noble, 1989; Smigel & Ross, 1970). It was found that while the reasons cited for rejection varied widely, the reasons for acceptance were consistent and typically more subjective: the most frequently cited reasons included 'interesting', 'significant', 'useful', 'good topic selection', and 'well-written'. 'Good writing presentation' appeared to be a crucial factor in both acceptance and rejection decisions.

Although the studies reviewed in this section varied in design, there is a lack of marked disciplinary difference for the core criteria of evaluation, contribution, writing, and novelty. By contrast, when justifying their decision of rejecting a paper, editors and reviewers tend to draw on subjective and subtle reasons, and the mostly cited reasons appear to vary by discipline. In order to gain better insight into such variations, this study has embraced the criteria for evaluation as one of its major targets of inquiry.

2.7 Rejection Rates

How likely is it that papers will be rejected? One group of studies examined the rate of rejection. Among these there were four cross-disciplinary studies (Beyer, 1978; Hargens, 1988; Miller & Serzan, 1984; Zuckerman & Merton, 1971). In the earliest of these, Zuckerman and Merton (1971) reported that humanity journals had the highest rejection rate, followed by social and behavioral sciences, mathematics and statistics; physical, chemical and biological sciences had the lowest rates, 'running to no more than a third of the rates found in the humanities' (p. 75).

Later studies showed similar results. In general, the rejection rate was much higher in 'soft' sciences than 'hard' sciences; the highest rate was observed in business, economics, management, and education, followed by social sciences and psychology, and the lowest rate was found in natural and physical sciences (Beyer, 1978; Gans & Shepherd, 1994; Hargens, 1988; Henson, 1988; Miller & Serzan, 1984).

The pattern of disciplinary variation in rejection rate corresponds to that of reviewer agreement. Some researchers attributed the variation to differences in the shared norms among editors, reviewers, and authors on research quality (Bakanic et al., 1987; Cole et al., 1988; Zuckerman & Merton, 1971); 'when scholars do not share conceptions, theoretical approaches or research techniques, they tend to view one another's as deficient and unworthy of publication' (Hargens, 1988, p. 147).

Another influential factor is discipline-specific principles of publication. Fields such as physics have a principle that a paper should be published unless it is wrong, medical fields take great care in publishing accurate results, and social science fields are inclined to reject a paper unless it makes a significant contribution (Bakanic et al., 1987; Beyer, 1978; Cole et al., 1988; Gordon, 1980; Zuckerman & Merton, 1971). Editors and reviewers who are aware of, and conditioned directly by, such principles would follow the same principles in reviewing, and the absence of a clear definition of 'significant contribution' in some fields was believed to relate to the markedly high rejection rates in those fields (Chubin & Hackett, 1990; Fölster, 1995; Mackie, 1998).

Space availability in journal issues is identified as another likely source of disciplinary variation in rejection rate (Beyer, 1978; Budd, 1988; Hargens, 1988; Zuckerman & Merton, 1971). The study by Beyer (1978) revealed that physical science journals published more issues each year than social sciences journals, and articles published in physical science journals were much shorter which allowed space for more papers to be accepted. The study by Hargens (1988) in the same field revealed that changes in the rejection rates were affected by changes in the number of submissions over time and 'structural differences between the scholarly communities' (p. 149) instead of space shortage.

In addition, a journal's prestige (usually ranked by impact factor) and its rejection rate were found to be closely related (Coe & Weinstock, 1983, 1984; Rotton et al., 1993; Weber & Stevenson, 1981). As Psathas (1976, p. 173) asserted, 'to be a prestigious journal… requires that it be exceedingly hard to get published in'.

Three studies directly compared the rejection rates of the same journals at different points in time (Coe & Weinstock, 1983; 1984; Hargens, 1988; Moyer & Crockett, 1976). By comparing the findings of the studies (Appendix 8), it also becomes clear that rejection rates have increased over time. Although the studies rarely explained causes for the increase, it was implied in the literature that this was partially due to the 'academic information explosion' after World War II (Moyer & Crockett, 1976, p. 489), with the growth of submissions outnumbering the growth of journals initially, and then later reflecting the increase in IT application and publishing pressure of recent years. Moyer and Crockett (1976) also found that in business and management the rise in rejection rate was more drastic for second- and lower-ranking journals than for top journals, and they argued that the 'academic information exploration has resulted in a relatively greater increase in the quality of the second-tier journals' (p. 492).

There were two exceptions to this general trend in the literature. Hargens (1988) recorded the rejection rates for 30 journals in three major fields in late 1960s and early 1980s and found that, although the rejection rates of most of the journals had increased, those of eight physical science journals had decreased. Via's study (1996) targeted the same library and information science journals as two earlier studies (Budd, 1988; O'Connor & Van Orden, 1978). By comparing the results of the three studies Via showed a decline in the rejection rates for a majority of the journals studied, and argued that the decrease was a result of the proliferation of new journals in that field.

Overall the pressures associated with peer review and their impact on editors and journals, reviewers and authors appear to be growing, and this is becoming clear in rejection figures, as well as in the debate on how reviewing should occur. In this context what is the fate of the rejected papers?

2.8 Fate of the Rejected Papers

Since a majority of initial submissions are rejected, what happens to the rejected papers? Studies devoted to this topic have generally identified that most authors try again to get their work published (see Appendix 9 for a more detailed breakdown of the information).

Gordon (1984) asked 98 biochemistry authors whether they would resubmit after a rejection; 59 percent said yes, and indicated that they would not resubmit it to a lower ranking journal. However, as some other studies revealed, the majority of rejected papers tend to be submitted and accepted by journals of lower status or impact factor (Daniel, 1993b; Liesegang et al., 2005; Lock, 1985; Nemery, 2001; Opthof et al., 2000; Ray et al., 2000) or smaller circulation (Abby et al., 1994; Chew, 1991; Daniel, 1993a).

There was also evidence that the longer a paper stayed unpublished, the more likely it would end up in a lower ranking journal (Ray et al., 2000). Only a few found evidence against this down-grading effect (Campanario, 1996a; Gans & Shepherd, 1994; Hernon et al., 1993; Weller, 1996).

Authors do not always agree with the final decision. They are normally given the option to appeal to the editor. Only one empirical study explored what happened to authors' appeals. Simon et al. (1986) examined 2,337 papers submitted to a leading sociology journal between 1977 and 1981 and found only 3 percent of the rejected authors appealed against reviewers or the final decision and that associate and full professors were more likely to appeal. In regard to editorial actions on the appeals, in 63 percent of cases the editor sent the paper to new reviewers, in 22 percent of cases the editor made an in-house decision, and for the rest no clear action was taken (mostly because the authors just 'expressed' themselves without requesting specific action). Twenty-six percent of the 74 decisions were changed as a result of the appeal (Simon et al., 1986).

With respect to resubmitted papers, some journals did not identify this status to reviewers but treated them as new submissions (Chew, 1991). If reviewers knew that

the paper was a resubmission, as Jauch and Wall (1989) revealed, they would still apply the same criteria as they would to a new submission, but would take into consideration the comments of previous reviewers and the authors' responses and expect the paper to address the previous reviews adequately.

The evidence reveals that, on the whole, a paper rejected by one journal is usually accepted by another eventually, often of a lower rank, after several resubmissions. It supports what was discussed in Chapter 1 about the functions of peer review, that it does not only *control* and *improve* quality, but also *stratifies* quality by 'matching' papers of certain quality with journals of certain ranks.

2.9 Efficiency of the Peer Review Process

The whole process of peer review costs time and effort and a balance has to be achieved between quality and speed. For example, after the initial submission authors will expect a series of waiting periods, including the time taken to receive an acknowledgement of receipt, a letter of the initial decision of outright rejection or sending out for review, and editor's decision letter accompanied by reviews; if revised, they have to wait for the final decision and if accepted to see the article in print. Several studies provided insight into time lags involved in these procedures. This information is tabulated in Appendix 10.

Some patterns of publication lag were evident. The lag was much longer in soft sciences than in hard sciences. Factors such as the use of multiple reviewers, multiple rounds of revisions, and extensive copy-editing were believed to contribute to the prolonged publication lag in the soft science fields (Beyer, 1978; Gordon, 1980). Authors would expect to wait longer if they submitted to a general journal than to a specialised journal (Ely et al., 2005; Mason et al., 1992; Yohe, 1980), or if they were lower in academic status (Zuckerman & Merton, 1971).

An earlier study in management revealed a trend of growth in publication lags (Moyer & Crockett, 1976). This trend was not atypical nor did it ease off with time. Ellison (2002) examined the editorial records of five top and 22 average economic

journals of the past three decades and revealed a dramatic slowdown in the publication process – at the top-ranking journals submission-acceptance time had increased by 12-18 months. Ellison offered three reasons: the initial response time by the editor after the in-house review was lengthened, more extensive revisions were required, and longer time was taken by authors to revise the paper; but he also noted that most of the slowdown remained unexplained by available evidence (Ellison, 2002, p. 947).

Several studies explored the time taken by reviewers to complete a review; the findings are presented in Appendix 11. Clearly, the time a paper actually stayed with reviewers was not very long, compared with the entire publication lag, and the time reviewers spent on each paper was considerably less, sometimes less than five percent of the entire lag (Loonen et al., 2005). As Hamermesh (1994) argued, most reviewers 'are in no way responsible for the slow editorial process' (p. 158).

'Publication backlog', or the period between a paper being accepted and published, also varied among journals (Hamermesh, 1994; Mason et al., 1992; Yohe, 1980). The backlog seemed to link to several factors, including the number of issues published in a year, the priority of the topic and the routine process of proofreading and printing. Publication backlog was often to exceed one year in some economic journals (Hamermesh, 1994), and some nursing editors reported a backlog of one month to three years, with an average of seven months (Thyer & Myers, 2003).

Is publication lag caused by a shortage of resources with the editor or reviewers? An earlier study of some linguistic journal editors found that the 'shortage of time is the single most pressing problem' of the responding editors (Lavelle, 1966, p. 7) and most of them believed they could work more efficiently if more financial support was available. Economist Hamermesh (1994), who was veteran editor and editorial board member of a number of journals, found that higher quality reviewers (measured by citation and experience) took longer to complete a review – about three weeks longer than average reviewers and 6.5 weeks longer than junior ones, which implied an extra 'time cost' for using high quality reviewers. Hamermesh called this 'the implicit price of quality in the market for referees' services' (p. 160).

Would authors perceive the 'time cost' the same way as editors? Little attempt has been made to explore this beyond three studies. In probably the earliest study on the topic (Brackbill & Korten, 1975), some psychology academics indicated publication lag was their top concern out of 22 areas in peer review needing improvement; the maximum publication lag they would accept at that time was 3.96 months.

Economics authors in Mason et al.'s study (1992) attributed the lengthening lag to editors (12%), reviewers (46%), or both (28%); some felt too many reviewers had been employed for each paper, and suggested that the editor screen all papers before sending them out, set up a deadline for reviewers, check their progress frequently, provide timely reviewers with financial rewards, and punish tardy reviewers probably by banning them from publishing in the journal; some suggested simultaneous submission be allowed.

In a study by Swan (1999), most of the 2,500 responding authors saw publication lag as a major barrier to the publication of their work; they were concerned that someone might publish very similar work elsewhere earlier than them, or their work might be dated by the time it was published.

Many authors responding to the survey by Nicholas et al. (2006) desired a 'faster publishing process in order to both have quicker access to the results of research and to have their own works published faster' (p. 197); they and many others (for example, those in Swan's survey) expected the use of electronic publishing to speed up the process.

The timely provision of reviews and editorial decisions is something over which 'reviewers and editors can exert direct control', according to Gilliland and Beckstein, (1996, p. 680), and it is a key to enhancing authors' perceptions of the quality of the editorial process. However, as shown in this section, publication lag is not caused solely by the actor, the environment, or the process; it is complex and thus is difficult to determine what and how much additional support will be needed to reduce lags.

2.10 Mechanisms to Improve Peer Review

Problems of peer review and the need to address them have been tackled in many major books, for example those by Lock (1985) in his book *A Different Balance: Editorial peer review in medicine* (pp. 108-132), Daniel (1993b) in *Guardians of Science: Fairness and reliability of peer review* (pp. 63-69), and by a group of editors in *Editors as Gatekeepers: Getting published in the social sciences* (Simon & Fyfe, 1994).

Academics' concerns coalesce on two major aspects – the practice of editors and reviewers and the efficiency of the process. Brackbill and Korten (1975) collected 22 suggestions for changing peer review from a group of psychologists and then sent the suggestions to another group of psychologists to rate their importance. The four highest-rating suggestions included to: take measures to 'insure speedier review of articles'; ask reviewers and editors to 'supplement criticisms with helpful suggestions'; 'emphasize more strongly the total significance of an article rather than simply its technical soundness' in the evaluation; and for reviewers to become 'familiar with articles or other material' that are cited in the work (p. 938).

A study of 314 social work academics revealed that the areas in which they mostly wished to see changes included the efficiency of the process, the responsibility of editors and editorial board members, the capability of reviewers, and the quality and fairness of the reviews (Sellers et al., 2006). Mason et al. (1992) asked 632 economists to consider some approaches to improving the efficiency of peer review. 'Paying reviewers' was supported by most, while the option of post-publication review by readers was rejected by most. The provision of incentive for timely and quality reviews was suggested by medicine authors responding to the study by Polak (1995).

In a leading medical journal, a working party was formed to discuss standards to be applied in the evaluation of health economics papers, and the resultant guidelines were published in the journal. Two years later, a study was conducted by the editors involved in the working party to evaluate the effect of the guidelines (Jefferson et al., 1998). It was found that they made no difference to the quality of submitted papers as authors took little notice of them; while the editors being surveyed felt that the guidelines were useful in the internal review process, but did not specify exactly how.

Several studies tested the reliability of some evaluation scales for reviewers to use for the purpose of improving the inter-rater reliability of certain journals. Without exception, these studies concluded that the scale they promoted was reliable and useful (Cho & Bero, 1994; Feurer et al., 1994; Hargens & Herting, 1990; Strayhorn et al., 1993; van Rooyen et al., 1999).

Two studies tested the effect of training reviewers on improving review quality but identified only limited effect (Callaham & Tercier, 2007b; Schroter et al., 2004). Hatch and Goodman (1998) tested the effect of the practice of providing reviewers with supplementary material (namely, related unpublished papers or abstracts) and revealed that the majority of the responding reviewers felt that the material was helpful or would be helpful; and one third reported that some specific comments in their reviews were directly affected by receipt of such material.

2.11 Conclusion: A Framework for Studying Journal Peer Review

This and the first chapter have highlighted that journal peer review is a multi-faceted phenomenon with a long and irregular history of development and research, covering many thousands of journals operating more or less as independent units to fulfill seven primary functions that serve in the production and dissemination of knowledge.

a) The function of quality control
b) The function of providing expert advice
c) The function of establishing credibility
d) The function of improving manuscript quality
e) The function of nurturing collegiality
f) The function defending academic autonomy
g) The function of stratifying publications in a field

There are visible or tangible elements involved in peer review, but much of the key activity is intangible, particularly judgment, which is not easily accessible because it is 'internal' academic work.

A leading scholar and editor in management (Bedeian, 2004) noted that peer review is inherently subjective and that 'intellectual advancement is not a result of establishing objective truths, but rather a product of social definition' in addition to a process of negotiation. Bedeian goes on to say that published work is 'inevitably a compromise between what its authors want to say, and the form in which they must say it, so as to be judged acceptable' by editors and reviewers. This is because judgments about knowledge are filtered through a personal lens, which 'alters individual referees' understanding and shapes their thinking in an idiosyncratic fashion' (pp. 199-201).

Until very recently empirical work in the field has not focused on the complexity, but has tended to isolate the focus to what variables might be changed to improve the external work of peer review, to better tailor it for use in applying impact algorithms, whereas at its core what review does is assess manuscripts. We know only snippets about quality judgments, for example, there is high agreement between reviewers when it comes to rejection, and in identifying a paper of high quality, but what does this consensus mean? Studies of journal peer review rarely go deeper.

For such a high-profile activity it is surprising that multiple studies confuse the issues. Thus we find none of the actors' characteristics seem to have any influence when it comes to the final recommendation or to the quality of the reviews. What does this tell us? At the very least it is necessary to recognize the need to examine in tandem:

- the framework in which reviewing activities take place, and
- its active engagement with review as a practice.

This research identifies peer review as a phenomenon that manifests as a wider dynamic of operations with five dimensions (Figure 2.2). The dimensions have been identified as pivotal in the literature over several decades. They are the persons involved (actors) and their characteristics, the environment they operate in as editors, reviewers and authors; the intent or 'functions' of peer review; the actual practices, and the criteria both specified for and used by reviewers. When one begins to focus on judgment in peer review this means the researcher has to take into account the five dimensions in the following way:

The dimensions comprise external structures such as journal policies, the broader expectations of peer review as identified in the journals, local and external environmental influences, the practices of disciplines and the characteristics of the actors.

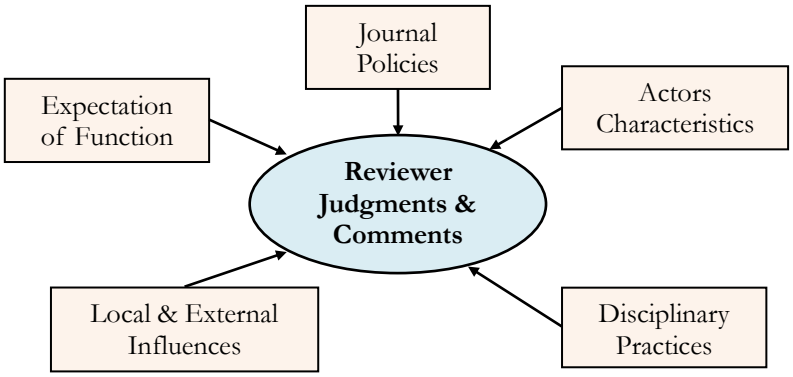


Figure 2.2 Dynamic of Peer Review Operations with Five Dimensions

Given the pressures now exerted on academe, peer review needs to be examined afresh and in the light of all the knowledge generated about it to see whether it is possible to address the concerns so often voiced, but which appear too difficult to deal with. There is a need to reject simplified views such as 'peer judgment is unreliable because there is a lack of consensus among reviewers' for a more complex conceptualization to accommodate its 'diverse stakeholders and the inconsistent demands they place on it' (Hackett, 1997, p. 55). The complex nature of peer review demands an extended, holistic consideration (Cain, 1999, p. 535).

Both Chapters 1 and 2 have identified some disciplinary variations in the perceptions of research assessment and in peer review practices, but few scholars have attempted to explore whether and in what ways the judgments of peer reviewers in different fields might differ, especially between social sciences such as education and hard sciences such as physics and chemistry – areas that have actually attracted very little attention in the peer review literature, but seen both as key building blocks to future knowledge and its operationalisation.

Yates, former chairperson of the Australian Association for Research in Education, has a section in her book, *What Does Good Education Research Look Like?,* devoted to peer review. She believes that the way research is perceived and assessed in education is fundamentally different from experimental sciences and it is 'a mistake' to treat these fields in the same way (Yates, 2004, pp. 25-27). This is not an uncommon sentiment in education and invites exploration. As the preceding review has shown, most of the studies focus on particular practices, in single disciplines, undertaken in isolated contexts without a sense of the inter-relatedness of the dimensions uncovered in this and the previous chapter.

This study has incorporated the above considerations to explore peer review across disciplines in a more holistic way, and specifically in a way to elicit the intangible practices of reviewers. As will be explained in Chapter 3, there are multiple phases and sources of data collected to address the following key questions:

i) How are the principles and practices of peer review defined within education and the physical sciences?

ii) How do reviewers perceive and address their role and activities in their respective disciplines?

iii) What judgments are made about quality and why, and do either differ by discipline?

One of the themes of the literature is the highly individualistic nature of reviewing practice and this is what is explored in more depth by this research alongside disciplinary difference. In addition the research seeks to offer a way of looking at peer review, but particularly the decisions behind the judgments, more holistically, to see if it is possible with depth of insight to identify novel solutions to some of the tough, seemingly unanswerable questions that remain in the field.

CHAPTER 3: RESEARCH DESIGN AND METHODS

3.0 Introduction

The journey through the peer review literature presented in the previous two chapters revealed the growth in research in the field, and a substantive discourse about peer review and its roles and functions, but also identified a number of gaps and a general lack of a critical mass of findings in some important areas. Some voices were calling for a more concerted effort to be undertaken to determine more about reviewer and editor judgment and it became evident that comparative studies involving two or more disciplines were atypical. Some other aspects about peer review that suggest new angles for study include:

- The long-term stability of its core functions
- The complexity of the phenomenon as a system of knowledge and activity
- The lack of detail on the more intangible and invisible activities of peer review and the need to draw these together with the visible tangible elements of reviewing and decision making.

Focusing on what have been perceived as very different fields of research, education and physics and chemistry, the research questions in this thesis address the above in whole or in part.

The questions driving the study were identified in chapter two and are articulated further below:

In education and physics and chemistry

1. How are the principles and practices of peer review defined by scholarly journal policies and do they differ from each other and other disciplines?
2. What is the impact of the wider environmental influences on publication and peer review?

3. What are the functions and expectations of peer review as perceived by those involved in reviewing?
4. What problems have beset peer review and how have they been addressed?
5. How do reviewers understand the task of reviewing and how do they approach the task in practice?
6. What criteria do reviewers emphasise or draw on in making judgments about the quality of a paper?
7. How do reviewers differentiate between papers in terms of quality?
8. How do reviewers communicate their judgments and justifications to authors?

These questions require *both* comparable quantifiable information and intensive meaning making that are the hallmarks of a mixed methods design.

This chapter will describe the research design and data collection methods, beginning with an explanation of the rationale for undertaking a mixed methods design. It outlines the multiple phases of data collection and details sampling strategies and instrumentation; discusses considerations of ethics and of validation, and concludes with an overview of the plan for data analysis.

3.1 Rationale for the Use of Mixed Methods Approach

The multi-dimensional nature of the research topic and research questions requires a mixed methods design. It is argued by methodologists that a combination of methods can provide access to a range and depth of information otherwise unavailable to a qualitative study or a quantitative study conducted separately (Flick, 2002).

Existing studies of journal peer review have by and large featured designs that are: discipline-dependent in approach; single-perspective or single-procedure in focus; and conducted primarily by 'journal-insiders'. The studies have spanned disciplines and have typically reflected the research traditions of the researcher. For example, researchers from medicine frequently conducted randomized controlled trials (for

example, Callaham et al., 2002; Cobo et al., 2007; Hirsch, 2004); those in business and economics often used regression analysis and developed models (for example, Beyer et al., 1995; Laband, 1987; Starbuck, 2005); and those in humanities and the social sciences were more likely to use qualitative approaches than those in the hard sciences.

The single facet nature of many studies underplays the complexities of the phenomenon being studied. For example, studies that tested the level of reviewer agreement typically did so by comparing pairs of reviewer ratings (for example, Cullen & Macaulay, 1992), but rarely did they conduct interview follow-up with the editors or reviewers to identify the reasons for the decision. Similarly, most of the studies of 'reasons for rejection' asked reviewers to list reasons that they would consider in reviewing (Glogoff, 1988) without interviewing them or analysing their reviews to see whether and how consistently the reasons cited were actually applied.

In terms of most of the studies being undertaken by 'journal-insiders', they have an obvious advantage that they can access, for example, all manuscripts submitted in certain years or information about the total pool of reviewers for the journal, or achieve a very high response rate to surveys, but the limitation is the results are journal-specific and there are issues with validity. There are possibilities for bias and reactivity. For example, if reviewers know they are being studied by the editor, they could be subject to 'reactive effects' (Webb et al., 1966). As one medical editor admits,

> I doubt that I or any of my fellow editors would casually maintain whatever biases we may have if we knew we were participating in a project aimed at testing those biases. I also doubt that most reviewers would carry out their customary behaviours if told that the behaviour would be under close scrutiny… the announcement of a test will surely alter customary behaviour. (Feinstein, 1991, p. 340)

Another feature in many peer review studies is the reliance on secondary data and descriptive analysis. In an analysis of the peer review literature, Weller (2001) summarized this feature as the 'lack of strict scientific methodology' (p. 320). Of the

385 studies identified in this study, about 70 percent undertook a retrospective analysis of editorial records (for example, information of authors, submissions, or reviewers, rejection rates, reviewer recommendations), or published records (for example, characteristics of published articles, citation rates). Thirty-three (8.6%) of the studies located drew upon reviews as the primary data source. Fourteen of these conducted systematic 'content analysis' (Silverman, 2006) of the reviews (these are discussed at the start of Chapter 2 and further in Chapter 8).

About 28 percent of the 385 studies involved the administration of a questionnaire to editors, reviewers, or authors. Eight studies employed open-ended interviews, and none of these targeted the reviewers about their perspectives of what they understood by peer review and how they engaged with the practice.

To sum up, in what is an otherwise active body of research, there is clear limitation in both practice and focus, which has the potential to distort the knowledge base of this important yet controversial area of academic life and research quality. By adopting a cross-disciplinary comparative, outsider approach, and with an intensive focus on reviewers as specified in the questions above, this mixed methods study aimed to address the identified limitations. As Tashakkori and Teddlie (2003) indicate, mixed methods have particular value when a researcher is trying to solve a problem that is situated in a complex context.

In order to inform the research design, the researcher sought to identify examples of mixed methods designs in the peer review literature. The conceptualization and utilization of mixed methods has advanced markedly in the past decade, especially since the launch of the *Journal of Mixed Methods Research* in 2007. One emphasis in the debate is about the extent to which the qualitative and quantitative methods should be integrated and what a 'genuine integration' should be like (Bryman, 2007).

Because most of the studies identified were conducted prior to 2000, earlier materials (Bryman, 1988; Tashakkori & Teddie, 1998) were drawn on to guide the search. Two 'loose' criteria were used: a study was included as 'mixed' if it used both qualitative and quantitative data collection and/or analysis methods and related

qualitative and quantitative data at some stage in the research process. On occasion it was difficult to determine because authors might publish the quantitative and qualitative sections of the same study separately and not refer to the 'mix' in those articles. For instance, Nicholas and colleagues conducted a questionnaire-based survey of some 5,000 authors worldwide. A large number of open-ended comments were collected and published separately from the quantitative data (Nicholas et al., 2005a; 2005b; Nicholas et al., 2006).

Five studies met the two criteria; two were conducted in medicine, one in biology, one in psychology, and the other in business management. Among them, two employed a sequential design, using data collected from the initial qualitative phase to inform the design of the subsequent quantitative phase.

Sternberg and Gordeeva (1996) explored 'What makes an article influential?' by asking 20 psychologists to list any features that made an article influential in the field. They then synthesised the responses into a questionnaire which they sent to a random sample of 500 psychologists for ranking the features on a 6-point Likert scale; the 'mix' of the two methods occurred when the questionnaire was constructed.

Street et al. (1998) examined the positive and negative behaviours of editors as perceived by authors. They interviewed two focus groups of management academics from which nine statements of editor behaviours were identified which, with another 11 statements generated from the literature, constituted a questionnaire. The questionnaire was then administrated randomly to 499 authors, who were asked to rank the statements on a 5-point Likert scale. Street et al. (1998, p. 3) called this design 'a two-stage integrated research strategy', but the integration happened only in the construction of the questionnaire.

Another two studies conducted a qualitative phase and a quantitative phase separately. Findings of the two phases were presented to explain the same issue from different angles, but the interpretation of each data set was also discussed separately (Murphy & Utts, 1994; van Rooyen et al., 1998b).

The purpose of Weller's study (1990a, 1990b) was to explore editorial practices and at the same time to examine the relative methodological strengths of questionnaire and interview data on this topic, so in effect it was strategically mixed but not a traditional mixed methods design. Weller claimed that 'initial interviews were very helpful in designing the questionnaire' (1990b, p. 263) and noted 'the value of interviewing a sample of questionnaire respondents as a means of clarifying and validating information obtained on the questionnaire' (1990b, p. 265).

Journal peer review is a complex, multi-dimensional process bound up with academic work, identity, disciplinary traditions, funding pressure and other systemic influence. Questions such as those identified for this study, and indeed for many of the studies identified in the literature, are deceptively simple but nonetheless many-layered. Hence a mixed methods design is both appropriate to the task and pertinent to advancing knowledge of the phenomenon.

3.2 Features of Mixed Methods Design

Johnson et al. (2007) asked 31 leading mixed methods methodologists how they define mixed methods. Five major themes were identified: what is mixed; when and where the mixing is carried out; breadth of mixing; reasons for mixing; and orientation of the research design. The orientation can be 'bottom-up' whereby the research questions drive the design, 'top-down' where the design is driven by the researcher's quest to conduct the research (pp. 118, 122-123), or possibly both within the same study. Section 3.1 has discussed the rationale for mixing for this study.

The current study adopted a general definition of mixed methods research provided by Johnson et al. (2007, p. 123):

> Mixed methods research is the type of research in which a researcher… combines elements of qualitative and quantitative approaches (for example, use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration… within a single study.

Each of the research questions outlined previously addresses a particular aspect of journal peer review; each of them embraces both a qualitative perspective and a quantitative perspective. This practice is called 'critical multiplism' by Cook (1985, p. 22) in an earlier article about social science research methods. As such, with regard to the themes concluded by Johnson et al. (2007), the orientation of this design is not dichotomously bottom-up or top-down, but situated somewhere along the 'bottom-up/top-down conceptualization continuum' (Johnson et al., 2007, p. 123).

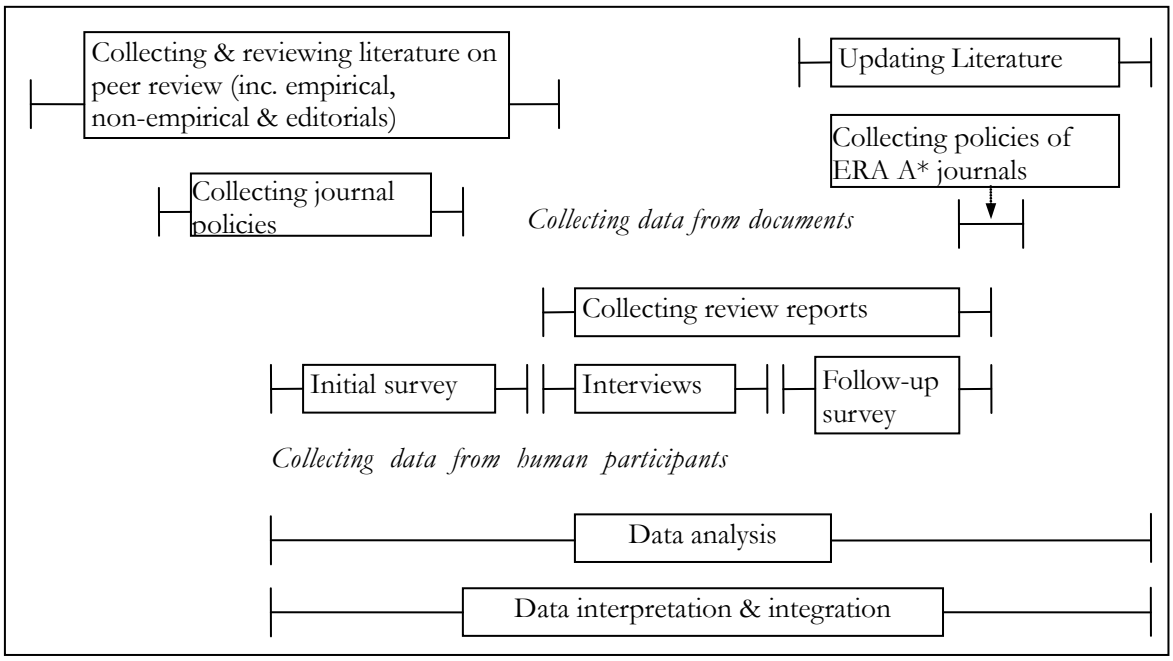
According to the methodologists cited in Johnson et al. (2007), 'mixing' can occur at all stages of a study, and involves not only research methods and data, but paradigms. Traditionally, qualitative paradigms and quantitative paradigms were deemed to be incompatible with each other. Alluding to difficulties in reconciling the two paradigms, some methodologists turned to the philosophy of pragmatism, stemming from the work of John Dewey, William James, and George Herbert Mead (Tashakkori & Creswell, 2007a), and established pragmatism as a 'third paradigm' (Johnson et al., 2007, p. 112).

Pragmatism is now regarded as the 'philosophical partner for the mixed methods approach' (Denscombe, 2008, p. 273), which provides opportunities for connecting qualitative and quantitative methods and gains knowledge 'in the pursuit of desired ends' (Morgan, 2007, p. 69). Arguing for a properly integrated methodology, Morgan maintains that the strength of pragmatism lies in 'its emphasis on the connection between epistemological concerns about the nature of the knowledge that we produce and technical concerns about the methods that we used to generate that knowledge' (p. 73).

In this study, the multiple phases of investigation involve both qualitative and quantitative approaches of data collection and analysis, and these phases have occurred at times both sequentially and concurrently. This is a dynamic process in which different phases are able to be mutually 'informative' (Greene, 2008; Tashakkori & Creswell, 2007b), especially in that the instrument of a later phase can be amended progressively as a result of the findings of earlier phases so as to ensure important and emerging points are captured. The next section explains this model.

3.3 Research Design for This Study

The data collection activity involved two aspects, one drawing on documents and the other on human participants, and six phases of investigation. Figure 3.1 shows all the activities associated with the project. The lines and boxed text show the rough timeframe of an activity in relation to the others. The horizontal arrangement of the boxes represents sequential occurrence and the vertical represents concurrent occurrence. Figure 3.2 shows the six phases of investigation. Detailed explanation is



presented in the following paragraphs.

Figure 3.1 Schematic of Data Collection Methods of the Multiple Phases

DATA & ANALYSIS

PHASES	**Documents**	**Human Participants**
	Phase I Analysis of selected editorials *Textual analysis*	Phase III Initial surveys of journal reviewers *Statistical and textual analysis*
	Phase II Analysis of journal policies *Statistical and textual analyses*	Phase IV Interviews with journal reviewers *Textual analysis*
	Phase VI Analysis of reviews	Phase V Follow-up surveys of journal reviewers

Textual analysis	*Statistical and textual analyses*

Figure 3.2 Data, Phases, and Methods of Analysis

In this study, physics and chemistry have been treated as one group representing physical sciences. This is because both physics and chemistry are laboratory-based fields; physics journals often embrace chemistry in their scope and vice versa, and the total population of senior academics in physics and chemistry is approximately the same as that of education.

3.3.1 Phase I: Analysis of selected editorials

Editorials are a form of commentary written by editors and published as a 'prelude' to virtually every issue of the journals they edit. They are a direct communication to readers and potential authors, and editors often take editorials as an opportunity to convey their concerns and expectations about review in general and their journal in particular. The initial literature search identified 174 editorials across disciplines that focused on journal peer review. They represented the perspectives of editors from a wide range of disciplines and types of journals.

The analysis of the editorials focused on three themes: how is peer review process implemented in scholarly journals, what are the criteria for evaluation suggested by editors, and what problems arise in peer review and what solutions if any are identified in the editorials? Results of the analysis are presented in Chapter 4.

3.3.2 Phase II: Analysis of journal publishing & reviewing policies

The second phase targeted peer review practices and evaluation criteria as stated in journal policies. This objective was pursued in two steps: an analysis of the journal policies of an extended sample of education journals, and a comparison between the policies of the top-ranking journals in education and in physics and chemistry.

Education as a field was chosen as the starting point for this phase of study. Although similar efforts are available in the literature (Altman & Schriger, 2005; Eldredge, 1997), they are all based in fields where well-established indices are used to classify and rank journals, gauged by impact factors. In education, by contrast, there is no one single and simple measure of journal quality; and impact factor has been

rejected by many education researchers as an adequate measure (Coleman, 2006). So the researcher began her work by becoming familiar with the policies of 700 journals (Lu, 2006). Further work was truncated by the publication of the introduction of Excellence in Research for Australia (ERA), where for the first time journals were 'assigned' to fields and ranked. The ARC released a list of ranked journals (ARC, 2008) in which 745 education journals, 331 physics journals and 375 chemistry journals are ranked into four tiers and this was used in the study of journal policy

The researcher collected and analysed the profiles and reviewing policies of the top ranking (5% at that time) education and physics and chemistry journals on the ERA list, working from the principle that these top journals would not only have well-established policies, but the quality of the policies was directly comparable and represented the pinnacle of expectations in the selection of quality papers. The analysis aimed to detect any fundamental disciplinary difference in the peer review practice and evaluation criteria of journals. This is reported in Chapter 5.

3.3.3 Phases III-V: Initial survey, interview, and follow-up survey

The other strand of this study draws on human participants – senior Australian academics who are also experienced reviewers. It consists of three steps: an initial mailed questionnaire survey (Phase III), open-ended telephone interviews (Phase IV), and a brief follow-up email survey (Phase V). The aim was to explore, from several angles, reviewers' individual understandings and expectations of reviewing in general and how they identify and distinguish quality in particular.

A stratified purposive sampling strategy was adopted for this phase. This strategy is a mix of features of probability sampling and purposive sampling. The two strategies are traditionally regarded as dichotomous: probability sampling is used primarily in quantitative studies with the purpose to achieve representativeness by sampling a large number of subjects (Tashakkori & Teddlie, 2003), while purposive sampling is used in qualitative studies to obtain greater depth of information from a smaller sample of carefully selected subjects who can best answer the researcher's questions (Maxwell, 1996; Patton, 2002). Stratified purposive sampling brings the two strategies together in a mixed methods context. As Teddlie and Yu (2007, p. 90) explain:

> The stratified nature of this sampling procedure is characteristic of probability sampling, whereas the small number of cases typically generated through it is characteristic of purposive sampling… the researcher first divides the group of interest into strata…and then selects a small number of cases to study intensively within each strata based on purposive sampling techniques. This allows the researcher to discover and describe in detail characteristics that are similar or different across the strata or subgroups.

This study targeted professors, associate professors and senior lecturers in 37 Australian universities (the researcher's own university was excluded to avoid bias) in the fields of education, physics and chemistry who had served as journal reviewers.

Table 3.1 Target Population (valid in March 2005)

	Professors		**A/Professors**		**Senior Lecturers**		**Total**
	Male	Female	Male	Female	Male	Female	
Education	60	29	71	62	115	133	930
Physics	53	5	59	6	66	8	
Chemistry	77	2	80	8	75	22	

By browsing university websites, a list of 930 academics was generated and their demographic information and postal addresses were recorded. By first following the 'probability sampling' approach, the researcher divided the list into strata by status, sex, institution, and then by discipline in an Excel spreadsheet. This resulted in 18 discipline-sex-status strata, as shown by the bolded group of cells in Table 3.1.

Table 3.2 Sample of the Initial Questionnaire Survey

	Professors		**A/Professors**		**Senior Lecturers**		**Total**
	Male	Female	Male	Female	Male	Female	
Education	15	7	18	15	28	33	232

Physics	13	1	15	2	16	2	
Chemistry	19	1	20	2	19	6	

A purposive sample was drawn from approximately one out of four names from each stratum, with a caution of keeping the balance of institutional representativeness. This process generated a sample of 232 individuals (Table 3.2).

After the preparation, the first round of invitations was dispatched in October 2005. In each invitation package, there were enclosed in that order:

- a cover letter explaining briefly the study and the content of the package;
- an information statement;
- a consent form;
- a cover letter to the pre-interview survey (which explained the letter might only be read if the person consents to participate);
- the initial questionnaire (Appendix 12);
- the interview protocol (Appendix 13);
- a draft sample of the follow-up survey; and
- two addressed, stamped envelopes (to enable participants to post the consent form and pre-interview survey and review reports separately if preferred)

Colour-coding was used to differentiate disciplines: questionnaires were printed on paper of three different colours: green for education, yellow for chemistry, and blue for physics. The contents were identical. Each questionnaire was assigned with a four-digit ID before it was sent. The first digit denoted discipline, 1 for education, 2 for chemistry, and 3 for physics. The second digit denoted sex, 1 for male and 2 for female. The third and fourth digits were 'streamline codes' assigned to each individual. For example, a male education academic Alan Andrew (pseudonym) would be coded 1101, and a female chemistry academic Anne Andrew (pseudonym) would be coded 2201.

The ID was clearly marked on the right top corner of the first page of the questionnaire. It and other material were then enclosed in an envelope with a name label that matched its ID. The returned consent forms, surveys and reviews were also marked with respondents' IDs, and their names were removed from all research

records. This process not only protected the confidentiality of the participants, it enabled in-depth analysis of data collected from the same source at different stages.

The respondents were first asked to complete a survey of four A4 pages in length, containing four sections of closed-format items with a fifth section for free comments. The questionnaire was designed in a simple format that aimed to describe 'the characteristics of a sample at one point in time' (Mertens, 2005, p. 172), that is, respondents were 'concurrent' in status and experience at that point. This approach suited the objectives of the study, namely, to gain detail that would be informative on its own, be comparable with previous studies in the field, and underpin and complement the interview to follow. The brevity also recognised the time pressure on the informants and was easy to complete. The questionnaire is provided in Appendix 12.

As Grant and Fine (1992) maintain, a questionnaire provides a means by which a researcher can obtain systematic data about a group. Across 40 items this survey drew on a variety of response types including Likert-scale based responses, yes-no answers, and data entry such as specific numbers of grants or publications. Five aspects of information were collected, including: background of the reviewers; their opinions of peer review as a method of research assessment in their field; their experiences as reviewers and authors; their level of success as a researcher in grant application and journal publication; and any additional comments that they had on peer review. The findings are reported in Chapter 5.

The initial survey was followed by in-depth interviews. The participants were asked at the end of the initial survey to nominate a time and date and a telephone number if they consented to participate in the interview. The participants who offered their consent were contacted to make a time. For those who did not consent, a thank-you email was sent to acknowledge their contribution to the initial survey.

The use of interview was based on the premise that the 'perspective of others is meaningful, knowable, and able to be made explicit' (Patton, 2002, p. 341) and for the purpose of interpreting 'phenomena in terms of the meaning people bring to

them' (Denzin & Lincoln, 2000, p. 2). The aim of the interview was to probe into the reviewers' understandings of quality and reviewing and therefore to 'validate the information in the questionnaires and to obtain more data than could be acquired through questionnaires alone' (Weller, 1990b, p. 262). As the participants were located all over Australia, the interviews were conducted via telephone. This method is recognised as a very convenient way of overcoming distances both in space and in time, and enabling the retrospect of past events or experiences (Perakyla, 2005).

The interview was constructed in a structured yet open-ended fashion. The protocol (Appendix 13) consisted of 13 questions carefully worded and arranged with the intention of taking each participant through the same sequence primarily to build a picture of how they distinguish between papers and to maximise the opportunity to scrutinise key topics and facilitate thematic coding of the transcripts (Flick, 2002). Before each interview, the protocol was reviewed to reflect new themes emerging from the initial survey or previous interviews; unanticipated topics that arose in the interviews were pursued if time permitted.

The interview protocol was trialled with volunteers from the researcher's own institution to last for about 30 minutes. The interviews were recorded by a digital recorder with the permission of the participants, and transcribed jointly by the researcher and a professional transcriber.

The interview protocol had three parts. The first six questions addressed the general expectation and practice of journal peer review; the next four questions were pertinent to the criteria for evaluation, and the last four questions asked the reviewers to compare journal peer review with grant peer review from their experience as ARC assessors or applicants. No other study has taken the opportunity to compare different types of peer review to highlight and elaborate the features of decision making in journal peer review. The responses are discussed in Chapters 6 and 7.

The interview transcripts were read alongside the findings of the initial survey and any interesting points of conjunction or convergence were noted. A follow-up

survey of three pairs of questions was generated from the notes, which probed further into topics that emerged from the responses to the initial survey or the interviews but were unanticipated in the initial design and those that indicated disagreement among the respondents. This survey was emailed to those who had participated in the initial survey and/or interviews. The findings are reported as the second part of Chapter 6.

3.3.4 Phase VI: Textual analysis of review reports

The last phase was the collection of reviews volunteered by the participants. In the initial invitation, the participants were asked if they would provide some reviews they recently prepared for journals. Due to ethics considerations they could not be matched with the respondents. Sixty reviews from 28 participants were collected. They were content-analysed and compared with the interview data. The categories used to code the interview transcripts were also used to code the reviews. The findings are reported in Chapter 8.

3.4 Ethical Concerns of the Study

Research ethics are a set of principles that guide the researcher to conduct an ethical study (Johnson & Christensen, 2002). In order to ascertain the appropriate conduct of the study, the researcher considered the possible range of ethical concerns and attended to them carefully. The project was approved by the Human Research Ethics Committee (HREC) of The University of Newcastle, Australia, in July 2005 (*Approval No H-056-0605*).

An information statement about the study was provided to all the individuals initially contacted; how and why they were selected, what their role would be, how information would be collected from them, and how anonymity and confidentiality would be maintained were explained. The participants were asked to sign a consent form. They could choose to participate in one or more phases of the study by ticking the appropriate boxes on the form. The decision to participate was entirely their choice. An interview or survey was only conducted, and reviews obtained, if the person was willing to participate. Contact details of the HREC were provided so

that the participants would be able to contact the HREC if they had any concerns or complaints about the study.

The researcher requested that participants de-identify any reviews before sending them so that at no time could she identify the author or the journal for which a review was prepared. In the Information Statement, the researcher suggests that participants provide reports only if it did not contravene any agreement or contract that they might have entered into with the journal. It was expected that the participants, all of whom were senior academics in their field, were capable of making an adequate judgment upon action.

Hardcopies of the consent forms, the questionnaires, the interview transcripts, and the reviews are stored in a locked filing cabinet in the researcher's office. The electronic records are stored in the researcher's computer secured with personal login name and password. Access to the data is restricted strictly to the researcher and her supervisor. Participants' identities on any of the research records were replaced with ID codes before the data analysis stage began. Although their demographic information was used in the data analysis and comments were quoted, they were reported in a way so not to identify any individual.

3.5 Mixed Methods Integration

The actual methods of analysis are reported in the relevant chapters, but as indicated above employ the use of *SPSS* for the quantitative data and *QSR NVivo6.0 software* for the textual data. This next section explains how the integration of methods and findings was conceptualised and operationalised in the study.

Integration of qualitative and quantitative findings occurred during data collection and through subsequent iterations of separate analysis and then combined analyses of the various sources of findings. This process extended the 'breadth of mixing' (Johnson et al., 2007, p. 122) to the degree of 'maximum integration' (Woolley, 2009, p. 19; Yin, 2006, p. 42).

The mixed methods design adopted by this study can be viewed as a continuum,

where the pattern of mix is shaped by the research intentions and starts with the research design, consolidated at the stage of data analysis, and finalized with the completion of the thesis. The researcher attempts to achieve 'genuine integration', so as to both ensure the soundness of this study and provide a sound model for similar studies in the future.

The notion guiding the plan of data analysis is provided by Bryman (2007, p. 21): 'In genuinely integrated studies, the quantitative and the qualitative findings will be mutually informative. They will talk to each other… to construct a negotiated account of what they mean together.' As Bryman notes, in contrast to triangulation, a 'genuine integration' not only tests the degree to which 'findings are mutually reinforcing or irreconcilable', but is 'about forging an overall or negotiated account of the findings that brings together both components of the conversation or debate.' That is the style of the final chapter.

As Bryman (2007, p. 21) indicates, apart from many other potential barriers to genuine integration, the major challenge for mixed methods researchers is 'to find ways of fashioning such accounts when we do not have established templates or even rules of thumb for doing so.' It was found that many mixed methods studies kept both the analysis and interpretation of different data sources separate, and rarely was the integrative approach undertaken in the stage of data analysis (Greene et al., 1989; O'Cathain et al., 2007). The researcher designed the study with the intention to achieve thorough integration from design to analysis and to interpretation.

In this study, the goal of 'genuine integration' was guided by the theory of O'Cathain et al. (2007, p. 150) that draws on four analytic strategies and an interpretative strategy that were suggested earlier by other methodologists, including Caracelli, Greene, Onwuegbuzie, Sandelowski, Tashakkori, and Teddlie. The five strategies are: extreme case analysis, data conversion, typology development, data consolidation, and crystallization of findings. The following paragraphs explain their application in this study.

Extreme case analysis refers to the process of identifying extreme cases, or outliers,

by a quantitative approach and then probing those cases by a qualitative approach; the qualitative data will provide exploratory insights of what makes a case an outlier, as compared to the other cases in the sample. In the two surveys, the analysis of journal policies and the initial survey, any extreme cases identified were noted and probed by analysing their 'parallel components' – the editorials and the interview transcripts, respectively.

Data conversion, also called 'data transformation' (Creswell, 2002, p. 220), refers to the practice of converting one type of data into the other to allow statistical analysis or thematic analysis of both types together. For example in addition to the NVivo files of the interviews and reviews which can be quantified in the software,

- The research literature was archived and categorized to improve navigation and continuous reference, and quantified where applicable
- The journal policies were archived, categorized and quantified to improve navigation for continual reference, and to determine the emphases
- The editorials were archived, categorized and where applicable counts were used to reveal emphases.

Such transformations facilitate comparison or 'mutual conversation' between the two types of data.

Typology development is where the analysis of one data type yields categories that are then used to analyse the other data type (O'Cathain et al., 2007, p. 150). In this study, this strategy was applied back and forth between data types. For example, certain themes that emerged in the analysis of journal policies were used to guide the analysis of editorials; themes identified in the interviews were used to form categories to analyse the reviews. This practice also helped 'eliminate or minimise key plausible alternative explanations for conclusions drawn from the research data' and 'elucidate the divergent aspects of a phenomenon' (Johnson & Christensen, 2002, p. 299).

Ultimately, the researcher utilising a process of continuous, systematic testing of emerging hypotheses and questions, came to a point where all the data were in verifiable patterns, and categories were saturated and cross-checked.

Data consolidation is where qualitative and quantitative data are jointly reviewed to create new variables or data sets. According to O'Cathain et al. (2007), this strategy is mainly applied to large quantitative data sets. For the current study, the data set was relatively small, so this strategy was applied tentatively. When the follow-up survey was constructed, findings from the initial survey and the interviews were jointly considered to develop questions to be asked in the follow-up survey. The purpose of this practice was to 'dig deep' into the reviewers' perceptions about some key issues of concern or points of relevance requiring more clarification.

Lastly, while the above four strategies are pertinent to data analysis, the strategy of crystallization of findings is more for the interpretation of findings. Sandelowski (1995) provides a classic discussion of crystallization. This strategy involves a practice in which findings from different components of a study are compared. In this study, the effect of crystallization was achieved by comparing findings generated by qualitative and quantitative methods that investigated the same issue, in order to identify any convergence, divergence, or discrepancy. The findings chapters, as a whole, attempt to complete the 'continuum of mixing'. The next chapter will move on to the discussion of findings generated from the analysis of the editorials.

CHAPTER 4: EDITORIAL PERCEPTIONS AND EXPECTATIONS OF PEER REVIEW

4.0 Introduction

The complexities in pinpointing what occurs in review – the visible and invisible work, as well as the nuances in purpose have long been acknowledged:

> Not only those outside the scientific establishment, but also many researchers and reviewers involved in peer review [and] even editors who oversee the process differ in their views of how it should work and what its purpose should be. (Relman & Angell, 1989, p. 827)

One particular group, namely, publishers developing journal databases have to work with this complexity in identifying peer review status (Eldredge, 1997). Moreover, it is not easy to determine the corpus of journals for any field. Coming up with such information is, according to one information and communication scientist, 'a tale that… needs to be appreciated' (Tenopir, 2004, p. 32).

The primary aim to explore the phenomenon of peer review required, in the first instance, that the researcher obtain a thorough sense of the scope of what editors and journals specify, particularly in the discipline fields of education and physics and chemistry. This chapter reports on the two phases of investigation devoted to this aim.

In the first phase, a collection of editorials discussing peer review were analysed to explore how editorial needs and expectations were communicated to readers and authors. In the second phase, the policies of top ranking journals in education, physics and chemistry were examined to determine how 'quality' and 'peer review' were defined in journals of recognised high quality as well as to explore whether there were variations in such policies by discipline.

This chapter reveals how the researcher went about studying public knowledge, namely, visible information about what constitutes the criteria for selecting articles

for a journal and the stated mechanisms behind the practices. It identifies the key themes and patterns in this public material. Work of this breadth has not been reported previously, and in this study it serves as an important backdrop to the disciplinary experiences of journal review reported in later chapters.

4.1 An Examination of Selected Editorials

Editors express their views in private correspondence with reviewers and authors and in editorials. The former is not easily accessible, but editorials are accessible and are numerous. A total of 176 editorials were collected in the initial literature search, published between 1967 and 2007 in journals from a wide range of disciplines. The disciplinary distribution of the editorials is presented in Table 4.1. Although most journals publish editorials regularly, medical editors appear to be most likely to use the opportunity to address the topic of peer review.

Table 4.1 Disciplinary Distribution of the Editorials

Discipline	**N**	**%**
Business & Economics	3	1.7%
Physics & Chemistry	11	5.7%
Education	10	5.7%
Science, Information & Mathematics	18	10.4%
Medicine & Biology	111	63.2%
Psychology	7	4.0%
Social science & Sociology	16	9.3%
Total number of editorials	176	100%

A preliminary analysis identified six recurring themes in the editorials. Table 4.2 presents the number of editorials that discussed each of the themes. Problems in peer review and in the publication process and journals' or editors' reaction to them received the most attention, followed by the description of individual journals' practices and criteria of evaluation. The following sections discuss them in depth.

Table 4.2 Number of Editorials in the Six Categories

Theme	**Count of Editorials** (%, total n=176)	
Practices of publication and reviewing	38	21.6%
Criteria used in manuscript evaluation	33	18.8%
Roles and responsibilities of editors and reviewers	20	11.4%
Problems in peer review	60	34.1%
Journals/editors' reactions to the problems	31	17.6%
Opening the process: Editorial debates and trials	10	5.7%

4.1.1 Practices of publishing and reviewing

Editors identified some unique practices of their journals, including how many reviewers they recruited to each paper, whether they used double-blind review, and how they dealt with reviewers, authors, acceptance and rejection.

In general, the editorial process begins after a paper is received. A number of papers will be rejected during a 'pre-screening' process. In many journals, the initial sorting is undertaken by editorial assistants assigned by the editor. The assistants have 'the most expertise in the subject of the paper' and 'a general familiarity with the most qualified people to serve as reviewers' (Babor et al., 1996, p. 1759). They read the papers, reject some and send the others to reviewers, read the returned reviews, and make a recommendation for the editor. The editor then reads the recommendation with the paper and the reviews, makes a final decision, and prepares a letter to send to the author. Many editors emphasized that while reviewers and editorial assistants play an important advisory role, the final decision rests with the editor.

Some editors noted that they felt bound by reviewers' judgments, especially 'when reviewers agree with each other', 'it did not matter' what the editor's 'personal judgment about this paper might have been' (Donmoyer, 1996, p. 23). When reviewers disagree, however, editors have to decide the appropriate course of action. Donmoyer (1996, p. 23) argued that such a decision 'should be based less on a counting of reviewers' responses than on consideration of the reviewers' reasons' and that a detailed case is to be preferred over 'cavalier' dismissal.

With respect to recruitment of reviewers, typically two to four were the norm; with more in a few cases (Richardson Jr., 2002). Editors indicated that they would use an extra reviewer 'when opinions differ markedly or the Editor is not satisfied with the review' (Yankauer, 1982, p. 207). In more recent years there appears to be a trend to use more reviewers for each paper, irrespective of discipline.

A number of editors explained how they recruited reviewers ranging from drawing on a database of reviewers from which they generate 'a list of potential reviewers with expertise in areas appropriate for the manuscript' (Baggs & Minicucci, 2000, p. 1), through to drawing on editorial board members, published authors, the reference list in the submitted paper or their own networks.

Editors seek the quality of 'competence' in a reviewer, but this is acknowledged as having 'many levels of meaning' by a sociology editor McCartney (1973b, p. 290), which is 'an obvious answer but one that says too little about a concept with many levels of meaning.' The 'multi-level meanings' can include expertise, availability, the likelihood of providing an unbiased and timely review, and a balance in expertise (for example, theoretical stance, statistics), potential conflicts of interest, country of origin, and institution (Babor et al., 1996; Dehmer, 1982).

Some editors indicated they evaluated review quality regularly and keep records for future reviewer selection. Green and Callaham (2006), editors of the *Annals of Emergency Medicine*, adopted a tiered system. In this system the editors rate each review at three levels: high, middle, and low, and every six months they evaluate reviewers by three performance criteria: 'measures of quality' (average review rating), 'productivity' (number of reviews completed), and 'reliability' (number of review invitations ignored or declined and number of reviews submitted late). Based on the evaluation, the editors 'move' reviewers up or down on the ranking and send further review invitations to the subset of 'better-performing reviewers' (p. 306).

This practice, and the general inclination of editors to use reviewers who are performing well, may lead to an overload of such reviewers. A medical journal's strategy to avoid this is to block a reviewer from selection for three months after assigning them to a paper (Estrada et al., 2006).

With regard to the practice of blind review, editors often cited the need to reduce bias and maintain fairness as a reason for using double-blind review. Double-blind review has been shown to be used more commonly in social science journals (Miller & Serzan, 1984) than the hard-sciences. In an editorial in the *Journal of Peace Research* Gleditsch (2002) offered some reasons for this. He states, 'the lack of widely accepted theoretical and methodological paradigms in the social sciences leaves them more exposed to partial and irrelevant judgments,' and also suggests that 'social scientists may also be more alert than natural scientists to issues of fairness and the social functions of evaluation systems' (p. 261). If comparable journals in a field are using a double-blind process then this is also cited as a motivation for using it (Gleditsch, 2002; Green & Callaham, 2006).

Editors identify difficulties in blinding authors' identities which extend to the need to 'purge a manuscript of textual or bibliographic references that might reveal an author's identity' (Hernon & Schwartz, 2003, p. 359). This practice can be helpful, but it can also be 'time-consuming' (Bligh, 1998, p, 569) and 'not mechanically feasible' (Yankauer, 1977, p. 136). It could be 'quite easy for an experienced reviewer to identify the author, particularly when an earlier version of the article has been presented at a major conference.' (Gleditsch, 2002, p. 259)

Differences in the practice of handling papers requiring major revision were also indicated in some editorials. On the one hand the revised paper is sent to the original reviewers (Stern, 1996); on the other the paper is treated as new submission and sent to new reviewers (Stull, 1989). Others review the revised paper internally by the editor or an editorial assistant. One Chemistry editor (Morrison 1985, p. 1505) referred to the pressure on reviewers as a reason for the latter practice:

> Experience has shown that some reviewers resent being asked for additional comments on a revised manuscript… In our efforts to serve our authors efficiently and to avoid overburdening our reviewers, only rarely do we go back to the original reviewers for a second review.

When a paper is accepted for publication, five journals, three in medicine, one in chemistry and one in library science noted that their journals would provide a copy

editing service to check references and style. As the literature reveals, the lack of editorial support for language and copy editing was one of the top concerns of authors (La forge & Coelho, 1998; Nicholas et al., 2005), and was seen to contribute substantially to the improvement of an article for its style and readability (Jean-Pierre et al., 1996; Wates & Campbell, 2007).

After a review is completed, further contact with reviewers is revealed to differ. Some editors send reviewers the decision and other reviewers' comments (Babor et al., 1996; Estrada, 2006; Morse, 1996) as a means to seal the relationship. Some provide incentives, for example, one sent reviewers an annual diary as a token of appreciation (Abdou, 2004), another recognised the best reviewers in an editorial (Estrada, 2006), and a third honoured 'top quality reviewers' by a special acknowledgement in the list of reviewers published annually (Green & Callaham, 2006). No monetary or other tangible reward was noted. Clearly, maintaining a group of competent, reliable and willing reviewers is important to editors.

4.1.2 Criteria used in manuscript evaluation

During pre-screening, some papers are 'returned to the author immediately with a polite explanation stating that both the journal and the author would be best served by abandoning the review process at this stage' (Babor et al., 1996, p. 1759). Reasons for the initial rejection cited by editors fell neatly into four groups, from the most common being: unsuitable for the scope or readership of the journal; duplicating other papers awaiting publication or already published in the journal and flawed in major ways, to inappropriately formatted (for example, ignoring the journal's guidelines for style).

Some editorials offered detailed explanations for the criteria, sometimes in the form of 'Guidelines for Reviewers', including 12 editorials by Bruce Squires, editor of the *Canadian Medical Association Journal*, as a series of 'What editors want from authors and peer reviewers' published between 1988 and 1991, as well as others in similar vein (Alspach, 1994; Bligh, 1998; Campion, 1993; Carnegie, 1975; Crawford, 1988; Laine & Mulrow, 2003; McCartney, 1973a; Siegelman, 1989).

Which criteria are priorised? A medical editor identified three core criteria which in his opinion 'characterize superior research papers', namely, *novelty, accuracy,* and *importance* (DeMaria, 2007, p. 1667). According to DeMaria, novelty is the 'the first and most important characteristic' that editors look for. There are four levels of novelty, the highest being that a paper is the first to report on a subject; the next is that a paper yields definitive data to resolve controversy, the third - extending previous findings, and the fourth that it replicates prior findings in a different setting. 'Accuracy' involves high quality conduct of research in all aspects, and 'importance' refers to the contribution of the finding; the latter two are relatively easier to achieve than novelty, and so are more expected by editors (DeMaria, 2007, p. 1667). It is worth noting here the distinction made between novelty and contribution which will be returned to at several points in this study.

4.1.3 Roles and responsibilities of editors and reviewers

Editors' expectations of reviewers did not vary obviously by discipline or from those revealed in the literature. Reviewers are responsible to four parties: the reader (reflecting the functions of 'quality control' and 'establishing credibility'), the author ('improving manuscript quality'), the editor ('providing expert advice'), and the wider research community ('nurturing collegiality' and 'defending scientific autonomy').

Reviewers' responsibility to readers is to select the best for publication where they 'act as intellectual guardians for the profession' to 'maintain standards of excellence for the journal', and 'their stance is disinterested' (Beaver et al., 1983, p. 326). It is also about ensuring the readers, who are often less expert than the authors or reviewers in the topic area, that what they read is worthy of reading (Relman, 1978).

For authors, reviewers are obliged to conform to all the requirements of ethical conduct and return reviews in a timely fashion (Lazarus, 1980; McCartney, 1973b; Price & Dake, 2002; Zellmer, 1977), and to help authors to improve their work. Some editors require reviewers to provide constructive feedback, which 'moves the manuscript closer to acceptance, a galley proof, and, most importantly, making a worthy contribution to the scientific literature in one's field' (Bearinger, 2006, p. 72). Providing constructive reviews 'requires more time and a higher level of detailed

feedback' but they are a worthwhile 'source of support and guidance for improving [authors'] presentation and writing skills' (Bearinger, 2006, p. 72).

Some editors noted a specific task of reviewers that, as experts who are up-to-date with the current developments in the field, they should identify literature overlooked by the author and identify submissions that duplicated those already published (Carnegie, 1975; Jacobsen, 1980; Lazarus, 1980; Rohrich, 2006).

If reviewers fulfill their dual responsibility to the readers and authors, it means they have also acted responsibly to the editors in terms of sharing their workload and broadening the knowledge on which editorial decisions, and suggestions for authors, are based (Alspach, 1994; Feinstein & Spitzer, 1989; Rohrich, 2006; Squires, 1987).

Journal editors, according to Lazarus (1980, p. 1527), act as 'expediter' and 'quality controller': they establish and publish standards for reviewing, identify good reviewers and remove bad ones, assure equitable and prompt reviews, make final decisions on publication, and ensure scientific integrity of the process. Other expectations of the role of editors include: mentoring new reviewers; encouraging prospective authors to write; initiating mechanisms to address problems in the process; and implementing high quality reviewing and publishing practice (Hanna, 1993; Jacobsen, 1980; Squires, 1987; 1991; Wackerle & Callaham, 1996).

Other roles of editors referred to in the editorials included 'to investigate and better understand the decision making and screening process' (Wackerle & Callaham, 1996, p. 77); to 'ensure that the strengths [of peer review] outweigh the weaknesses' (Squires, 1991, p. 89); and 'in the midst of the turbulent adjustment between the print and electronic cultures, to assert the indispensability of editing' and to show that editing enhances the author/reviewer communication (Graham, 1997, p. 60).

Overall, and disappointingly, editorials themselves do not give much more depth of insight, or greater visibility into, the work of the editor, namely, 'in the spectrum of creative functions between the act of authorship and the act of reading, the work of the editor is the least visible' (Graham, 1997, p. 60).

4.1.4 Problems in journal peer review

Problems in peer review discussed in the editorials, such as reviewer disagreement, bias and publication lag, are well-documented in the literature. Instead of reiterating the problems, the researcher focused on identifying potential common causes of the isolated problems, as well as any feasible solutions to them.

In the editorials, editors attributed a number of problems in peer review to the level of paradigmatic consensus in their fields. As the literature shows, 'soft' science fields are characterized by a lack of consensus about what quality scholarship should look like across the spectrum of sub-fields and approaches. One social work editor indicated, 'what is the 'best' has to be viewed as relative to what is possible and to what is thought to be 'good' by one's peers' (Meyer, 1983, p. 3). This can put editors in a difficult situation of matching reviewers to authors and reconciling disagreement between reviewers.

Donmoyer (1996) shared an instance in his experience as editor for *Educational Researcher*: four reviewers reading one paper, between them, ticked all five boxes on the evaluation form, from 'publish' to 'reject', and one of them ticked two boxes. Donmoyer linked this to the diversity of perspectives in the field, and posed two questions: 'whether a match-by-paradigm should be pursued?' and 'whether editors should select only reviewers with precisely the same paradigm as the author?' Indicating that he did not 'want to answer it prematurely', he called for an alternative principle to guide reviewing 'that neither thoughtlessly dismisses nontraditional forms of scholarship and scholarly discourse as 'incoherent nonsense,' nor treats it with 'indifferent superficial tolerance" (p. 22). He introduced three policies for his journal: encouraging submission of non-traditional work; employing reviewers of a wide range of specialties to assess such works; and if such a work is judged to be high quality by at least one reviewer, the editor should make a decision in support of it.

To identify reviewers for an impartial assessment of non-traditional work, according to science editor Glen (1989, p. 398), 'remains a difficult and important question' for editors. Editors who have made an effort to invite truly paradigm-threatening ideas

are in the minority; the difficulty is partly due to the fact that new theories are 'most likely to challenge ruling paradigms, and thus contravene ideas that potential referees have already subscribed to in print.' Donmoyer's (1996) proposal above is one such effort; another is posited by McCartney (1973b, p. 287) of the *Sociological Quarterly* who 'seeks the names of younger scholars' to 'balance some degree of breadth with a sensitivity to innovativeness.'

There are also problems in peer review that are related to scientific integrity and ethics. Knoll (1990, p. 1332) once compared peer review to a bureaucratic system which 'copes best with routine, when nothing much is at stake', but what is known about peer review suggests the system is vulnerable to many individual and organizational influences, and problems such as conflicts of interest, bias, and scientific misconduct are well documented in the literature.

Conflicts of interest, also called 'competing interests' by two medical editors (Abbasi, 2006; Hoey, 2004), rarely surfaced in the editorials. According to some, conflicts of interest out of personal or financial relationships are easily identifiable and can be declared, while those that arise out of competition, intellectual favourism or jealousy are less detectable, especially when reviewers are anonymous (Davidoff et al., 2001).

Bias can exist in the editorial process. 'An editorial policy is biased', according to Feinstein (1991), 'if it depends on doctrinaire, ideologic, or political beliefs about science, rather than on the scientific quality of the research', and the editorial process is biased 'if approval is affected not by a work's contents', but on the basis of 'the authors, institution, or auspices responsible for the research' (p. 339). Feinstein claimed his journal as 'the first and only journal in ancient or contemporary medical history to have an utterly unbiased review process' since the knowledge that the journal was 'uniquely unbiased' was 'a prerequisite for becoming an editor' (p. 339).

With regard to scientific misconduct, two editorials, one from social work and the other from public health, summarised misconduct on the part of authors, which included fraud, plagiarism, misinterpretation of data, fragmentation of data, multiple submissions, and gift coauthorship (Hopps, 1989; Susser & Yankauer, 1993).

Reviewer misconduct was also addressed by some editorials; among the problems are unduly harsh critiques or unfair rejection of a manuscript for venomous reasons, and plagiarism of authors' ideas (Baines, 1987).

One issue about the conduct of editors is whether they can overrule the editorial decision of an associate editor or acting editor or even their own, in particular when the decision is 'acceptance for publication'. McCarty (2002), a psychology editor maintained that they had the legitimate authority to do so. This editor and others provided cases of editors overruling previous decisions for reasons such as: '[having] accepted too many manuscripts for all of them to be published in a timely fashion' (the case of the *BMJ*, cited in Price & Dake, 2002, p. 195); 'outside political (legal) pressure' (the case of a health education journal, cited in Price & Dake, 2002, pp. 195-196), and a need to 'uphold the highest standards of quality' (McCarty, 2002, p. 201).

Another major cause of problems in peer review is the explosion of submissions. 'Submission explosion' has been experienced by almost all disciplines. In the 1970s, sociology editor Glenn (1979, p. 785) claimed that 'the established procedure for evaluating manuscripts perhaps could work rather well when submissions were fewer and when almost all of the refereeing was done by members of editorial boards' a framework that in itself brought 'a kind of accountability' especially as the editor could devote 'a great deal of time and attention to each decision'.

Far from this ideal, Neuhauser (1997) summarised the two direct consequences of submission explosion as intensified tension between editors and authors and a heavier demand on reviewers: due to the proliferation of submissions, the concern for editors now becomes 'how to reject more articles of top quality' given 'most editors accept more than they should in terms of quality and also quantity'. In addition, the line between excellence and exclusiveness has become 'blurred and controversial' (Goldwyn, 2005, p. 243). As a result, editors often have to use vague reasons to turn down some technically competent papers (Neuhauser, 1997).

The submission boom has also placed a greater burden on reviewers, who normally serve on a voluntary basis – 'There are a limited number of willing, rapid, good

reviewers who can be exhausted by too many requests for their help' (Neuhauser, 1997, p. 301).

4.1.5 Editors' reactions to the problems

In the editorials, no editor claimed that their journal was free from problems but also indicated that journals were tightening up their policies regarding fairness and integrity in the practices of journals and reviewers. Editors identified a number of what were becoming regarded as 'good' practices, specifically to:

- require authors, reviewers, and editors to declare potential conflicts of interest and publish such declarations with the article (Hoey, 2004; Davidoff et al., 2001)
- reject a manuscript at editorial level 'if there is a clear and substantial conflict of interest' existing with the author (Alvermann & Reinking, 2004, p. 11)
- state explicitly in the letter to reviewers that 'We expect reviewers to protect the confidentiality of the material presented' and to 'ensure that the enclosed manuscript is not disseminated or exploited', even to the extent that if a paper is discussed with a colleague, the particulars need to be 'specified in a letter to the editor.' (Marshall, 1995, p. 1913)
- require a signed statement by all authors to clarify their individual contribution to the submission (Babor et al., 1996)
- introduce open peer review, where authors and reviewers know each other's identities (for example, Alpert, 2007)

In 1990 David Sharp (p. 1391), editor of *The Lancet,* provided a summary of potential means for making the reviewing process fairer, including:

- double blind review
- open review by removing anonymity
- seeking reviewers' declarations of integrity or giving them behavioural guidelines
- rendering all approaches to reviewers formal and completely neutral

- proceeding to peer review via questionnaire or checklist rather than by more unstructured means
- increasing the number of reviews for one or two to many more
- sending every paper out of the office for review
- checks on fraud
- author appeal mechanism.

However well-intentioned, they were not adopted in full by the journal. One of the key reasons given at the time was that this degree of 'bureaucratisation' would bring about a loss of integrity and status in other ways (Sharp, 1990, p. 1391):

> All these correctives could be introduced by all journals but they represent further bureaucratization of science, and implicit in many of them is a lack of trust in the ability or integrity of the reviewer. Do we want to go down that path? Science is too bureaucratic already… a journal would rightly come under pressure to take every possible corrective action – but at a price. For instance the individuality of journals (and some of the pleasures of editing) would be replaced by uniformity and bureaucracy.

The examination of editorials reveals that problems discussed in the 1970s are still worrying editors of the 2000s. For example, the debate on the openness of peer review, namely, whether reviewing should be all blind, all open, or half-and-half, has been carried on for an inordinately long period of time with no consensus to date even among editors from the same discipline. It would seem changes to peer review continue to be difficult to make.

Glenn (1979, p. 785) once asked the question 'why have there not been stronger pressures for reform?' and the explanations he offered were that:

> … it is easier to criticize the present system than to suggest workable alternatives. And any suggestion for innovation is always countered by the charge that 'it won't work' – usually in the absence of any empirical evidence concerning its workability.

Second, in Glenn's field (sociology) those 'best able to promote change (the more prominent and better established ones) are not very motivated to do so', while the young and untenured ones were struggling to survive and so 'unlikely' to 'demand reform' (p. 785). One could extrapolate that this dynamic is still in place in all disciplines. Glenn concluded that 'there is much grumbling about the deficiencies of the system, and much hostility is focused on specific editors, but there is no effective pressure for change' (1979, p. 785). Editors are among the 'best able to promote change' in peer review yet are little 'motivated' to do so, especially when they find it difficult to collect empirical evidence to support such change (see for example, Feinstein, 1991; Sharp, 1990; Wessely, 1996).

The question is not whether or not to change, but how to change. Some editors, by claiming that they appraise their reviewing system regularly, show a preference for continuous improvement over re-engineering. According to former *BMJ* editor Richard Smith (1999b), re-engineering a process means to examine it and experiment with doing it in a totally different way. He believed that 'new entrants to the process of peer review may find ways to re-engineer it in ways that the old timers may find hard to imagine.' (p. 251). New technologies may offer such possibilities.

Continuous improvement as a procedure involves 'defining your processes in detail, collecting data on how they function, reflecting on how they might be improved, making a change, collecting more data to see if the process is improved, and doing this continuously.' (Smith, 1999b, p. 251)

The practice of some journals has reflected this idea to some extent. For example, the editorial group of the *BMJ* analyzed the practice of training reviewers of its own and comparable journals (Callaham et al., 1998; Schroter et al., 2004) and introduced a new reviewer training package with the aim to improve review quality (Schroter & Groves, 2004). Another persistent problem, evident in the lack of follow-up evaluation in this case, is the general lack of reported evaluation of any implemented change.

4.1.6 Opening the process

The call to 'open the process' has been a theme in editors' writings since the 1970s, and it has been heightened by the introduction of information technology, from emails and web-based submission to review management software. Advocates tend to promote openness as a panacea to cure many stubborn problems in peer review.

As observed in the literature, 'opening the process' refers to at least five practices, including: releasing reviewers' identity to authors; releasing reviewers' identity by publishing their names with the article; publishing both reviewer names and reviews with the article; post-publication reviewing, that is, after a paper has passed traditional peer review and been published and readers can post comments on a webpage; and 'open publication', an idea whereby virtually everyone is free to upload articles onto a journal website, and where the level of editorial intervention is minimized.

Figure 4.1 illustrates this 'Spectrum of Secrecy-openness'. The line in the middle divides the Spectrum into the left side of practices pertinent to identification of reviews and authors and the right side about the handling of reviews. The top-down arrangement of boxes illustrates the practices from secrecy to openness. The practices along the spectrum make up two distinct models: pre-publication reviewing and post-publication reviewing. The practice involving the release of reviewer identities in a pre-publication reviewing process is usually called 'open peer review', and post-publication reviewing is usually called 'open access'.

There is also a mixed use of the two models. An example is available from the *Medical Journal of Australia*: its editor conducted an experiment: papers that were already accepted for publication were posted online with the reviews, readers were invited to comment on the papers, and authors could revise their work in response to the comments before it appeared in print (Bingham, 1999). In the following discussion, the term 'open' is referred to the first model, open peer review.

Among the earlier participants in the debate were two physics editors. Manheim (1973, pp. 534-535) noted a moral pressure for reviewers to sign if they were given the option that 'choosing anonymity implies that there must be some reason behind

the choice' and concluded that 'the disadvantages of referee identification seriously outweigh the benefits'. Robertson (1976, p. 410) argued that the idea of open peer review 'has been brushed aside by most editors' as they 'fear it would lead to a number of serious consequences' and also 'it is easier to run a journal where exchanges between authors and referees can be controlled and if need be subdued, not because it is necessarily science that benefits from such anonymity'.

Before long, the idea of signing reviews was raised again in an editorial published in a sociology journal (Freese, 1979). The editorial was followed by comments made by other editors or researchers, published in the same issue. They all agreed that 'signed reviews almost certainly will create a different set of problems' (Bohrnstedt, in Freese, 1979, p. 244).

In 1977, the *Journal of Laboratory and Clinical Medicine* started to encourage its reviewers to sign reviews (Knox, 1981). This is the earliest trial of open peer review identified in the literature. After four years of implementation, Knox analysed the publication records and found that, in 1980, 60 percent of the reviews of accepted papers and 50 percent of the reviews of rejected ones were signed. Readers' opinions on the signed review policy were also varied. Knox (1981, p. 1) cited a reader's comment that, 'As an author – delighted, as a reviewer – a little less enthusiastic.' Knox concluded that the signed reviews were as valuable as the unsigned ones and decided to continue the practice in the interest of openness of communication.

Drummond Rennie, deputy editor of the *Journal of the American Medical Association,* claimed that the anonymous review system is 'a perfect example of privilege and power being dislocated from accountability' whereas signing reviews 'strengthens the link between power and accountability'. He asserted that to enjoy the benefits of open peer review, 'all that has to be done is for editors to agree to make it their journal's practice.' (Rennie, 1998, pp. 300-301)

'For editors to agree' however has appeared to be the main barrier to an open system (e.g., Rennie had failed to convince fellow editors in BMJ to unmask reviewers). A majority of the editorials conveyed a conservative attitude toward open peer review. Editors perpetually weighed its strengths against its weaknesses, but rarely made a

decision for their journal. They encouraged readers to comment on the options but at the same time chose to wait for 'stronger evidence for an open approach'.

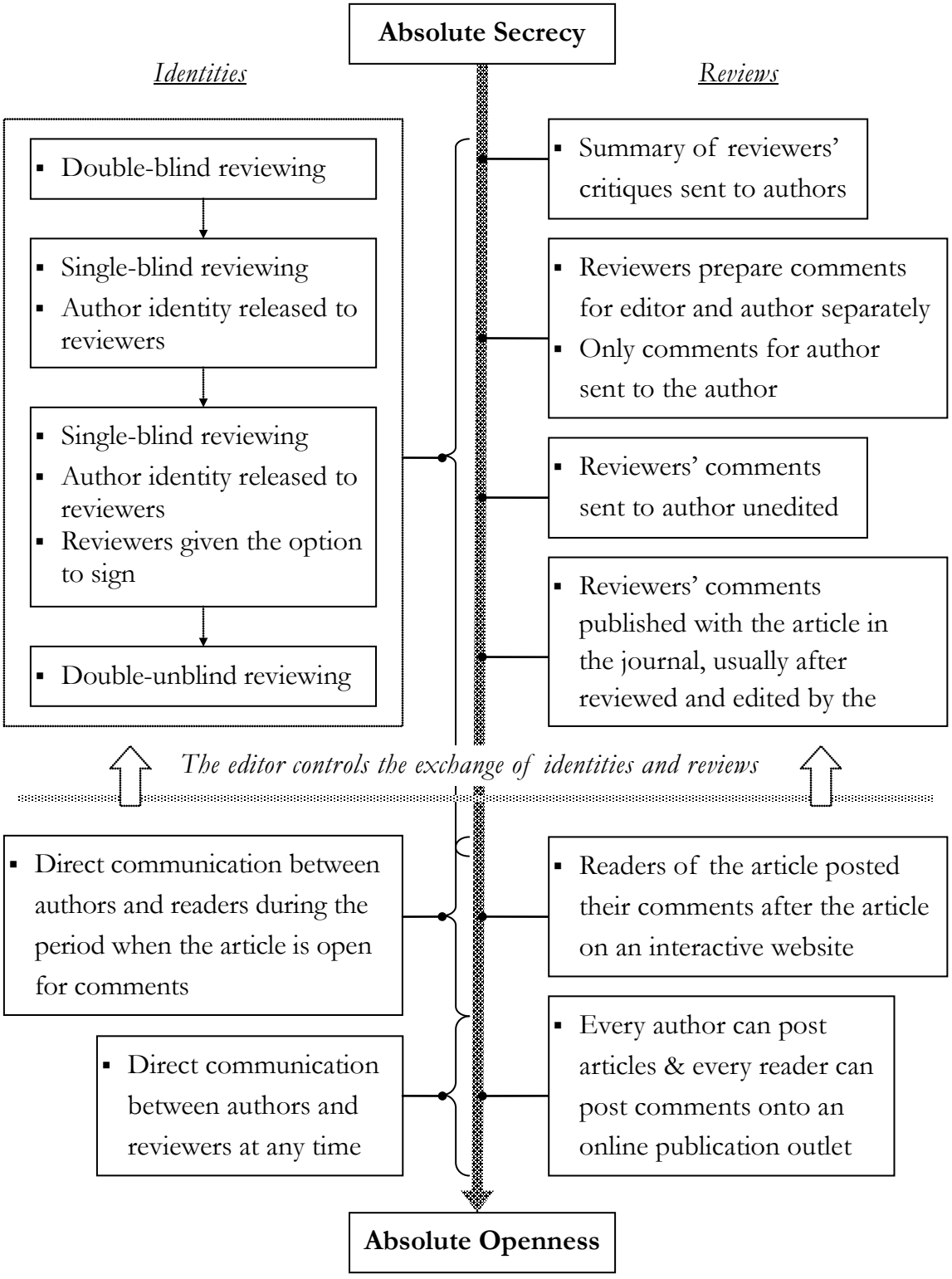


Figure 4.1 Spectrum of Secrecy-Openness for the Publication Process

Since the late 1990s, a number of medical journals have introduced the signed review policy, including the *BMJ* (Smith, 1999a), journals produced by BioMed Central (Godlee, 2002), *Journal of the Royal Society of Medicine* (Abbasi, 2006), and *PloS Medicine*. Godlee (2002, p. 2763) reported that,

> Authors like it; some reviewers decline to review openly but others welcome it; authors can identify conflicts of interest that reviewers failed to declare and editors are not in a position to detect; no adverse effects (such as careers ruined or reviewers beaten up) have been reported… signed reviews seem more constructive in their tone; and while it may sometimes be harder to find reviewers, any increase in editorial work is balanced by the time saved in not having to edit reviewers' comments.

Richard Smith of *BMJ* noted that, 'The sky hasn't come down'; no more than 30 of the *BMJ*'s 5000 or so reviewers quit after the policy, and the signed reviews did not become less critical (cited in Enserink, 2001, p. 2188). The *BMJ* went one step further by posting all signed reviews on its website with the article and, according to Smith (2006a, p. 181), 'had no effect on the quality of the opinion.'

Williamson (2003, p. 18) of the BMJ Publishing Group revealed that there was dissent among editors within the Group and that of the BMJ specialist titles, 'none has opted totally for open review. About two-thirds have given reviewers the opportunity to sign their reviews, and about two-thirds of those reviewers do'. He further reported that there were 'real misgivings, particularly in the smaller specialties'.

Arguably, to motivate journals to open their reviewing process, clearer and stronger external pressure needs to be present (for example, 'peak' bodies of scholarly publications in a discipline). It can be hard for outsiders to impose such pressure. Given the importance and unsettledness of the issue, it was explored further in this study with a group of experienced reviewers.

4.2 An Examination of Journal Profiles and Policies

It is always necessary for the researcher to become familiar with the way journals state their policies to get the complementary perspective for the editorials. When the researcher began this project in the early 2000s, there was no way to determine how journals were ranked in education so she began the process of mapping policies across more than 1200 journals and published on this (Lu, 2006). The initial work was useful in depicting the contemporary landscape of journal publication in education and identifying the types of information that journals shared with readers.

In the interim a decision was made in Australia to undertake a ranking of all journals in all fields for the research assessment exercise to be undertaken by the government (Australian Research Council, 2008). This made the earlier mapping redundant as the researcher could access the rankings when published. I then took the opportunity to compare the policies of top ranking education and physics and chemistry journals on this list. The researcher worked from the principle that the top journals would not only have well-established policies, but that they represented the pinnacle of expectations in the selection of quality papers. The analysis also aimed to test for any fundamental disciplinary difference in the peer review practice and evaluation criteria.

The Australian Research Council (ARC) released the initial Excellence in Research for Australia (ERA) list of ranked journals in 2008. This study examined the journal profiles and policies of all the top ranking journals in that list in education (n=45), physics (22) and chemistry (24). On the list, three journals are included for both physics and chemistry (*Journal of Chemical Physics, Nature Materials*, and *Surface Science Reports*); these, and another seven, are classified as 'physical chemistry'. As such, physics and chemistry journals were treated as one group in the analysis, which created two groups roughly equal in size for direct comparison.

Information of seven types from the published information of the journals was recorded, including: number of issues published yearly; commencement year; publishing body; country of origin of editors; sex of the editors; peer review policy; and criteria for evaluation. These types provide some sense of the disciplinary context in which reviewers operate.

4.2.1 Profiles of the top ranking journals in education, physics and chemistry

All the journals publish regularly, while physics and chemistry journals produce many more issues: 24 of them publish between 20 and 52 issues a year, and a physics journal publishes 90 issues yearly. There are four journals issuing once a year, only because they are '*Annual Review*' journals. The average number of issues published yearly by the physics and chemistry journals is 23.5, almost four times more than that of education (5.8 issues yearly).

Table 4.3 presents information about the publishers of the journals. It shows that a much higher proportion of the physics and chemistry journals are produced by learned societies alone or jointly than their education counterparts. Also, only major commercial publishing companies are involved in the production of top ranking journals in the three fields.

Table 4.3 Publishers of Top Ranking Journals in the Target Fields

Publisher	**Education** N=45		**Physics & Chemistry** N=43	
University alone	6	13.3%	1	2.3%
Learned society alone	8	17.8%	20	46.5%
Commercial publisher alone	15	33.3%	20	46.5%
University & learned society	0	-	0	-
University & commercial	2	4.4%	0	-
Learned society & commercial	14	31.1%	2	4.7%
University, society & commercial	0	-	0	-

The commencing decades of the journals are compared in Figure 4.2. As it shows, most of the top ranking physics and chemistry journals commenced around 1950s and 1960s, which was a decade earlier than the education journals. Although on average they were five years 'older' than the education journals, they included a higher proportion of journals commencing after 1985, as the extremities of the two lines show.

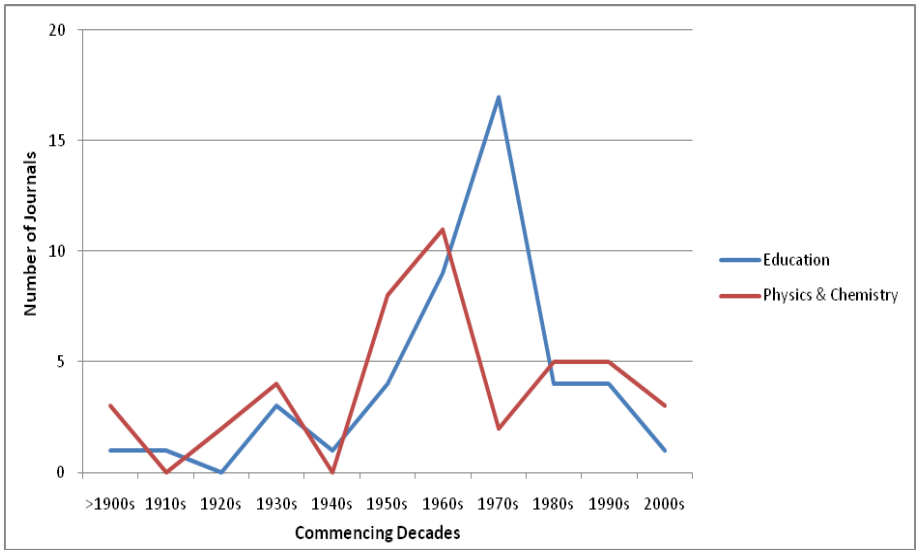


Figure 4.2 Commencing Decades of Top Ranking Journals in the Target Fields

With regard to editors, 27 of the education journals use a single editor, 14 have two co-editors, three have three co-editors, and one had four; among these there are 36 males and 33 females. By contrast, editor's posts in the top ranking physics and chemistry journals are clearly dominated by males. Of the 43 journals, 35 have one editor and 11 have two or more co-editors, of whom 95 percent are male.

About 42 percent of the education editors are based in the USA, 37 percent are in the UK, and the others are elsewhere. The geographic locations of the physics and chemistry editors are spread more widely. For example, more non-English speaking countries are involved in publishing internationally prestigious physics and chemistry journals. The USA dominates both fields and employed 54 percent of the physics and chemistry editors, which is a higher percentage than in education; the UK employed 8.8 percent; others included other European countries (20.6%, with Germany taking the lead with 10%), Asian countries (11.8%, of which 5.3% in Japan) and Australia (2.9%).

To sum up, a comparison of top ranking education journals and physics and chemistry journals reveals that the latter group produce four times more issues per annum, are generally 'older' (with a peak in the 1960s compared to the 1970s and steadier in growth overall), and learned societies play a more pronounced role in

their production. The physics and chemistry journals are published out of a broader range of countries, including some non-English speaking countries, while the publication of the education journals is clearly dominated by the USA and the UK.

4.2.2 Expectations of 'quality' in top ranking journals

All the journals publish detailed instructions for authors for the preparation of manuscripts, including types of articles, length, format of figures and tables, and referencing and layout style.

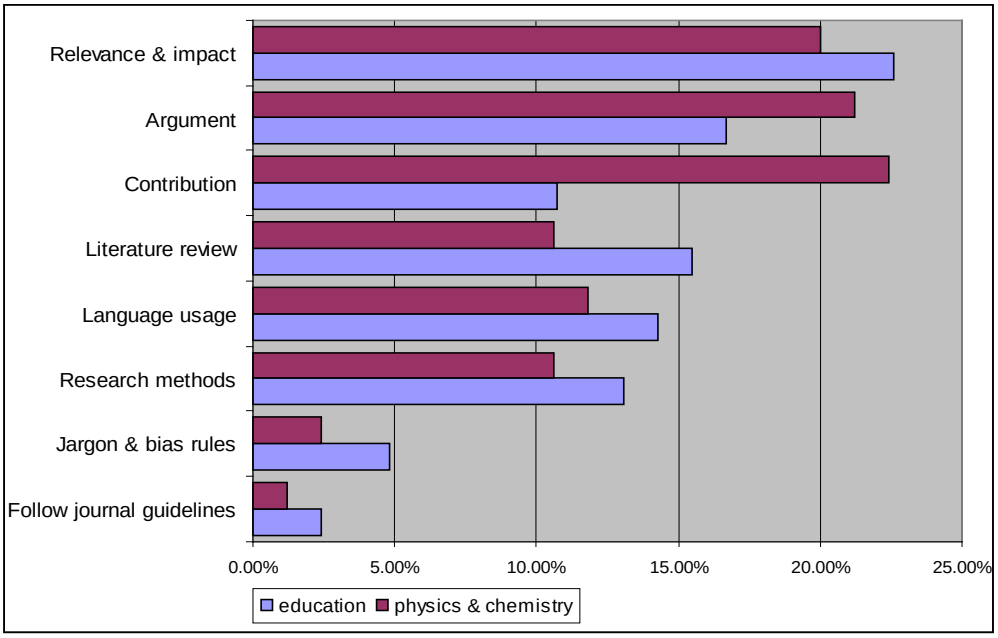


Figure 4.3 Proportions of Journals Specifying Identified Criteria

Half of the journals also specify criteria for evaluation that they expect authors or reviewers to follow, which are often presented as bullet-points. Very few journals explain the criteria in detail. The analysis of the policies identified eight broad groups of criteria (Figure 4.3). To get a sense of journals' emphases in their expectations, the policies were quantified on the basis of the criteria being present in any way or not. One journal was counted only once for each group. As Figure 4.3 shows, the top six criteria were noted much more frequently than the last two.

'Relevance and impact' is the most mentioned criterion. In the policies, the policies on 'relevance and impact' contain three aspects of requirements – suitability for the

journal, relevance to the journal's readership, and potential impact to the field the journal serves. The physics and chemistry journals clearly emphasise 'impact to the field' over 'relevance to the readership'; they often cite the former by using the words 'impact', 'the field', and 'the science community'. In contrast, the education journals clearly prioritise the 'relevance to the readership'.

To be a highly relevant paper, journals seek that a manuscript should:

- identify the populations that would most benefit from its publication and develop the practical implications for them
- draw out explicitly in a discussion of the research those issues or themes which are likely to have international relevance
- have as broad appeal as possible to the diverse readership
- stimulate debate
- be interesting to researchers in other related disciplines.

For the 'impact of the research', the journals, mostly from physics and chemistry, expect findings reported in a paper to 'be of interest and importance to the research community' and that it shows 'a possible impact on the ongoing research'.

In the policies the requirements of 'argument' emphasise four features, namely, clarity and conciseness, conclusion being substantiated by evidence, critical thinking, and sufficient detail being provided for the understanding or reproduction of the research. A paper marked by high-quality argument is one in which,

- material is presented in a clear, concise, and logical-structured way; the research approaches (for example, sample size, instrument, experiment, calculation) are clearly documented, accurate, and sufficiently complete
- the reporting is analytical not merely descriptive
- the data or evidence are substantial to support the interpretations and conclusions; unnecessary details or speculative ideas are avoided
- the discussion is written in a way that is easy to read and understand.

Requirements for 'literature review' expect both the quantity and quality of articles being quoted to be 'appropriate', in that they are 'relevant to the research topic and findings', 'the documentation of the literature should be selective rather than exhaustive', 'the implications should be clearly presented', and the quotes need to be 'referenced accurately'. Very few of the journals directly define 'a good literature review'. Instead, most of them focus on the functions that they expect the literature review to fulfil. This implies that the quality of a literature review is indicated by the successful fulfilment of its functions, including 'placing the work in context' and 'demonstrating authors' understanding of previously published literature in the topic' and whether they have 'given proper credit to related work'.

The two fields differ in the expectation of literature review's function of 'placing the work in context'. The education journals emphasise 'making sense' – situating and supporting 'the issue, the research problem, or the author's assertions' with the literature. For physics and chemistry, it is more about 'contribution' – to provide sufficient detail and background through the literature review for readers to better understand the significance of the work.

Requirements for writing and language usage refer to two different things in the policies. 'Quality of writing' involves similar requirements to those for the clarity and conciseness of 'argument'; both are concerned with the authors' ability to transform ideas into words. 'Language usage' indicates the expectation of using appropriate academic or scientific English, spelling, style and grammar.

Requirements for 'research methods' contain mainly broad statements, for example, a paper should display 'methodological soundness', 'scholarly rigour', 'scientific merit', and 'appropriate design'. Some journals identify specific requirements, which vary in focus. For example, one education journal requires that 'the guiding theoretical framework should be explained and justified'; another notes that the 'demonstration of internal and external validity is mandatory'; a third expects a paper to 'employ an established and recognized scholarly approach' and prefers 'qualitative research reports that employ sophisticated research designs and qualitative research reports that rigorously follow naturalistic research methods.'

Another two require the method to 'be clearly outlined and match the research question or the stated purpose.'

Comparatively, physics and chemistry journals put forward more objective, technical-oriented requirements for research methods. For example, one requires that 'the utility of new methods or techniques' and that 'routine experiments or calculations that simply extend previous methods to a new system are not appropriate unless the results are used to significantly advance the solution of an important problem.' Another requires 'adequate characterization being provided for representative compounds with regard to identity and purity.' Several others require 'nomenclature, mathematical formulae, symbols, abbreviations, and units' are 'correctly defined and used', and 'hazardous procedures' are 'clearly defined.'

Figure 4.3 shows that there are a much higher proportion of physics and chemistry journals publish requirements for 'contribution' in their policies than education journals. All the requirements are constructed very vaguely. Most of these present statements like, 'the paper should make significant/important/new/original/timely contributions to the field' or 'The prime criteria utilized to judge the quality of contributions will be their originality, significance, and novelty'.

In addition, four education and two chemistry journals note 'jargon and bias rules' in the preparation of manuscripts. The jargon and bias rules required that,

- the language and concepts that are standard in one subfield may be unfamiliar to non-specialists, also because for many readers English is a foreign language
- authors clearly explain the terms when their use is unavoidable
- national colloquialisms and idiomatic use of language should be avoided, and
- word choices and sentence constructions that might imply bias against persons on the basis of gender, racial or ethnic group membership, disability, sexual orientation, or age should be avoided

4.2.3 Peer review policies published by top ranking journals

Forty-six of the journals publish policy statements about peer review. The statements create a text file of 1,126 lines in total; most consist of one to several paragraphs, ranging from one sentence (in an education journal) to more than 200 lines (in two physics journals). On average, the statements from education are much shorter than those from physics and chemistry (10 compared to 45 lines).

The length of policy appears to have something to do with the type of publishers. Longer statements are more likely to come from journals produced by learned societies. Of the 16 statements exceeding 20 lines, 13 are from journals produced by learned societies; the other three are from journals produced by commercial publishers. The content of the statements were coded selectively (that is, not 100% coded) into four aspects as presented in Table 4.4.

Table 4.4 Journal Policy Statements on Peer Review

Issue of concern	Education		Physics & Chemistry	
	Line Count (total=273)	Journal (N=27)	Line Count (total=853)	Journal (N=19)
Reviewing procedures	190 (69.6%)	25 (92.6%)	285 (33.4%)	18 (94.7%)
Selecting reviewers	54 (19.8%)	14 (51.9%)	227 (26.6%)	15 (78.9%)
Anonymous/blind review	47 (17.2%)	13 (48.1%)	30 (3.5%)	5 (26.3%)
Contribution of peer review	2 (0.7%)	1 (3.7%)	52 (6.1%)	3 (15.8%)

As Table 4.4 shows, most of the policies were devoted to explaining the reviewing process, and more so in education than in physics and chemistry. By contrast how reviewers are selected is more prominent in the latter. The processes adopted by the journals are very similar. The full-fledged process consists of eight steps:

1) Logging and acknowledgement of the receipt of a manuscript;
2) Pre-screening the manuscript;
3) Selecting reviewers and sending them the manuscript;
4) Assessing the returned reviews and making an initial decision;
5) Notifying the authors of the decision and requirements for revision;

6) Assessing the revised manuscript, sometimes by consulting the original reviewers;
7) Making the final decision of publication; and
8) Handling author appeals.

The practices vary from journal to journal, for example, some skip one or more steps and some manipulate a step to different degrees. The variation is typically affected by the size, type, and tradition of a journal.

The literature shows that education journals tend to use more reviewers per paper than physical science journals. The current data confirm this and detect also the practice of many physics and chemistry journals of inviting authors to suggest potential reviewers and those to be excluded to 'help the editors identify reviewers with appropriate expertise'. The journals note that they generally honour authors' requests, although they are 'not obliged to select a name from that list'. This stated practice was not found in any of the education journals examined.

Also in many of the physics and chemistry journals to initially send a paper to only one reviewer, in fewer cases two, is common. Time is clearly the driving force in this practice. Some journals note that extended reviews are time-consuming and some state the purpose of reviewing is not to change an otherwise unacceptable paper into an acceptable publication. The policies indicate that editors may use an additional reviewer 'only in exceptional situations', such as when they feel the need for further advice about the statistics or techniques used in a paper, the initial reviewer is late in returning review, or if an author appealed against the final decision of publication.

The policies of nine journals spanning the disciplines offer detail regarding the selection of reviewers: they are chosen on the basis of expertise, familiarity with the subject of the paper, availability, free of conflict of interest, and the editor's previous experience of the person, such as the quality of his or her reviews and promptness.

In regard to anonymous/blind review, it was revealed that education journals more often use the terms 'blind' and 'anonymous' interchangeably in their policies while 'anonymous' is typically used by physics and chemistry journals; the latter are more

likely to adopt single-blind review and some even discourage reviewers to release their identities. As one physics journal stated,

> Unless they feel strongly… we prefer that reviewers should remain anonymous throughout the review process and beyond. Before revealing their identities, reviewers should consider the possibility that they may be asked to comment on the criticisms of other reviewers and on further revisions of the manuscript; identified reviewers may find it more difficult to be objective in such circumstances… If they wish to reveal their identities while the manuscript is under consideration, this should be done via the editor.

A number of the editorials also discuss the functions of peer review. For example, a unique function of journal peer review has been recognised, that it brings the wider research community together – it is 'a human enterprise we must all rely upon one another, as authors and referees, to do our best'; 'the journal reader benefits in that at least one independent expert has judged the paper to be new and interesting, to contribute to the advancement of the field, and to be without apparent flaws'; and 'the author benefits from feedback regarding the research and style of presentation as well as from pride in a refereed publication.'

What is evident in the statements is editors' public faith in peer review as 'the most workable' quality control mechanism. They also have faith in its contribution to scholarly publication – peer review conducted 'by independent, anonymous referees' is 'one of the most important reasons for the existence of a scientific journal'; it 'improves the manuscripts our journals publish'; and it 'helps to protect the literature, promote good science and select the best.'

4.3 Conclusion

Editors have many opportunities to publish their expectations and concerns about peer review, especially in journal policies and editorials. However, since editors are often the investigators of peer review research, but are seldom the subject of the research, these perceptions are largely unavailable in the empirical literature.

This analysis has shown that the foci evident in the two public sources of information are not very different from the existing empirical literature on peer review, namely, the focus is on its practices and problems. Publishing and reviewing practices of the top ranking journals in education, physics and chemistry, as stated in their policies, are very similar. The only obvious differences are that, compared to education, physics and chemistry journals publish more issues per year, use fewer reviewers per paper, are produced by a wider range of countries than the UK and the USA, and emphasise publication speed.

In addition, spanning disciplines editors tend to attribute most problems in peer review to the internal and external influences upon the system, including submission explosion, scientific integrity and the level of paradigmatic consensus in their fields. Also explored was the debate on open peer review, and a Spectrum of Secrecy-openness was developed to capture different editorial practices in this regard.

The analysis also identifies some less known or less tangible aspects, such as the expectations of reviewer competencies, a strong sense of faith in peer review, and a hesitation to change. The next phase of study further explores these perceptions with the other key group of actors in peer review – the reviewers, through in-depth interviews bracketed by two surveys. Chapter 5 presents the findings of the initial survey sent to a group of experienced Australian reviewers.

CHAPTER 5: SURVEY PARTICIPANTS' INVOLVEMENT IN, AND PERCEPTIONS, OF PEER REVIEW

5.0 Introduction

Despite the large number of academics who act as journal peer reviewers, relatively few empirical studies of peer review have concentrated on the specific activities of that group. In discussions about the academy the role and functions of the academic 'peer' are broadly fleshed out, especially in relation to building the disciplines, maintaining standards, preserving academic autonomy, and in the formation of scholars and academic communities. In Australia, investigation of national competitive grant review is rare, and the statistical work by Jayasinghe et al. (2001) in this field did indicate that those most likely to obtain grants (and therefore most likely to be very familiar with grant peer review) are professorial level researchers.

Exactly how peers execute their role as assessors has only recently become the subject of intensive research, notably Lamont's (2009) work on 'evaluative cultures' in the field of grant review which specifically examines the social dynamics involved in group decision-making about 'excellence'. Here she talks about how academics reach consensus on criteria and explores how norms are cognitively, socially and emotionally constructed.

It has also been revealed in the literature that because editors are often the ones writing about peer review (either doing research about the topic or commenting in editorials) that the deeper levels of their own understandings are rarely visible, nor explored in connection with their own experiences as reviewers and authors. The need for a more insightful, internal, dialogic process of individuals assessing what they do and why, and from different perspectives (for example, editor-reviewer-author) is raised by Graue (2006) in an article in *Educational Researcher* alongside her interest in achieving more from peer review through adding value, providing leadership, and developing clear statements of criteria for evaluation.

A feature missing in the peer review literature is the recognition that it is but one form of peer evaluation. The researcher hypothesised that the experience of acting formally in a number of roles as 'peer' may impact on the perceptions of the effectiveness and understanding of peer review. Asking reviewers to reflect on differences in reviewing in different contexts may also assist the researcher to identify more precisely what distinguishes certain practices of journal peer review noted in the literature, such as improvement the quality of manuscripts.

In the initial survey for this study nearly half of the participating reviewers were also editors. This chapter begins by summarising the questions asked of participants and then presents the findings from the initial survey.

5.1 Construction and Administration of the Questionnaire

When constructing the questionnaire, the researcher paid especial attention to framing questions that recognised existing knowledge, and pushed further to expand the boundaries of that knowledge.

The questionnaire contains 44 questions, divided into five sections (see Appendix 12 for a copy of the questionnaire). The first section collected the background information of the participants, including age, sex, education qualification, academic title, and field of research interest. It also looked at the participants' experience of serving as journal reviewers and ARC grant assessors and their reviewing currency and frequency, that is, the last time they reviewed a paper or a project, the number of papers or projects that they reviewed yearly and in total, their experience of serving as institutional research committee members and supervising research students. It was hypothesised that the nature, length and intensity in such research assessment-related experiences would have some impact on the participants' perceptions of peer review and the way they go about reviewing.

The second section invited the participants to offer opinions on peer review as a research assessment method in their field by rating various aspects on a six-point Likert scale. With regard to the effectiveness of peer review, previous studies often compared the before- and after-publication versions of a paper to see whether its

quality was improved by going through the peer review process; some invited authors to share their impression of the effectiveness of peer review (for example, Biddle & Aker, 1996). It was found that many authors believed the 'net effect' of reviewing was improving manuscript quality (Bedeian, 2003). What these studies had not attended to was whether 'improving quality' is among the ultimate considerations that reviewers bring to the task. This section included four items asking the participants to what extent peer review should aim to, and can, improve the quality of journal submissions and grant applications. Since the work of journal peer reviewers is autonomous and somewhat 'separated', unlike that of grant panelists, this section also asked whether the participants thought editors or the decisions grant committees followed/reflected their recommendations, and whether they brought different expectations to journal and grant peer review.

The third section focused on several specific experiences of the participants, including their initial motivations to serve as a journal reviewer or a grant assessor, whether and why they had declined to review certain papers or for certain journals, how often journals provided guidelines to them, how closely they followed these guidelines, and how useful they thought these guidelines were. This section also included an item to determine whether the participating reviewers gave different weight to journal peer review and grant peer review in terms of their importance.

The fourth section examined the participants' level of activeness and success as a researcher in grant application and journal publication. This section appeared at the end of the questionnaire in order to reduce resistance to responding. It invited the participants to provide the number of papers or projects they had submitted in recent years and the number of papers being accepted for publication and projects being funded. It was hypothesised that their level of success in research assessment would influence their views of peer review.

The last section invited additional comments. It was hoped that some interesting yet unanticipated issues would arise from the reviewers' open-ended comments.

The questionnaire was administrated to a sample of experienced reviewers employed by Australian universities, as discussed in Chapter 3. Usable responses were received

from 84 reviewers from 27 universities, or a response rate of 36.2 percent (of 232 individuals to whom the invitation was sent). Pearson's chi-square was calculated to test for systematic non-response bias by comparing the number of respondents to that of non-respondents by discipline. There was no evidence of non-response bias as the result $\chi^2(1)=.72$, $p=.337$ was not statistically significant.

The background information of the reviewers, including age, sex, academic title, and experience in research and reviewing, was analysed first to generate groups for further analysis. This was followed by the analysis of the ratings on items about the reviewers' perceptions. Frequencies, Chi-square, bivariate correlations (Kendall's *Tau*), and *t*-test were calculated. The open-ended comments were content-analysed.

5.2 Profile of Participating Reviewers

A total of 25 males and 22 females from education, and 34 males and 3 females from physics and chemistry, responded to the survey. Table 5.1 presents a breakdown of the respondents' demographic information grouped by discipline. As it shows, most of the respondents were experienced researchers and reviewers. All of the 84 respondents had a PhD degree; 37 (44%) had achieved their doctorate more than 20 years ago. There were 26 professors, 30 associate professors and 28 senior lecturers. Thirty-seven (44%) of them were from the 'Group of 8' – eight prestigious universities in Australia, recognised especially for research outcomes and research incomes, for example, government funding. No respondent was younger than 35; 28 (33.3%) of them were between 35 and 49, and 56 (66.7%) were 50 or older.

In order to detect any disciplinary difference, Cramer's *V* test was conducted for the four factors in Table 5.1. Cramer's *V* is a test used for a 2×2 table to measure the strength of nominal association. Statistically significant differences were found for all the four factors. Compared with those from physics and chemistry, the education respondents were much more senior in age while less senior in academic title; a greater proportion of the education respondents were from universities outside the Group of 8; and physics and chemistry were clearly male dominated, especially at senior academic level.

Table 5.1 Demographic Information of the Survey Respondents

Demographic Information		Education N=47		Phy&Chem N=37		Total N=84	
Age	Under 50	8	17.0%	20	54.1%	28	33.3%
	50 or over	39	83.0%	17	45.9%	56	66.7%
Gender	Male	25	53.2%	34	91.9%	59	70.2%
	Female	22	46.8%	3	8.1%	25	29.8%
Academic title	Professor	10	21.3%	16	43.2%	26	31.0%
	Associate Professor	16	34.0%	14	37.8%	30	35.7%
	Senior Lecturer	21	44.7%	7	18.9%	28	33.3%
Institution	'Group of 8'	16	34.0%	21	56.8%	37	44.0%
	Other universities	31	66.0%	16	43.2%	47	56.0%

Table 5.2 presents data on the respondents' experience with research assessment related activities. Although the sampling strategy did not particularly target academics who were both reviewers and editors, nearly half of them had served as editor for peer-reviewed journals. All the respondents had served as journal reviewers for four years or longer, and 61 (72.6%) had reviewed for more than 10 years.

With regard to the frequency and currency of reviewing, they reviewed an average of 14.2 papers per year, and more than 90 percent of them had undertaken a review within three months before completing the survey.

Table 5.2 Respondent Experience with Research Assessment Related Activities

Experiences	Education (N=47)		Phy&Chem (N=37)		Total (N=84)	
Editors of peer-reviewed journal	25	53.2%	14	37.8%	39	46.4%
Editorial board members	35	74.5%	22	59.5%	57	67.9%
Members of ARC panels	4	8.5%	3	8.1%	7	8.3%
External assessors for the ARC	15	31.9%	29	78.4%	44	52.4%
Assessors of other funding bodies	9	19.1%	13	35.1%	22	26.2%
On internal research committees	34	72.3%	27	73.0%	61	72.6%

Chi-square tests were conducted for the items included in Table 5.2. Compared to education, a lower proportion of the physics and chemistry respondents had served as editorial board members (χ^2=6.27, p=.012), while a higher proportion of them had served as assessors for the ARC (χ^2=20.00, p<.001) and other external funding bodies (χ^2=7.78, p=.005); the differences were statistically significant.

In addition, seven respondents served on the assessment panels of the Australian Research Council (ARC), Australia's major national funding body for basic and applied research in all disciplines except clinical medicine and dentistry; 44 (52.4%) had served as external assessors for the ARC, and 14 (16.7%) had done so for more than 10 years. They assessed an average of 7.6 grant applications per year. Also, 22 (26.2%) respondents had served on the assessment panel of other research funding bodies, and 61 (72.6%) had been members of institutional research committees.

The data also suggested that doing one type of research assessment related activity would bring opportunities of being invited to do more such activities. Kendall's *tau* coefficient tests were calculated on the experiences of the respondents. Kendall's *tau* is a non-parametric hypothesis test which measures the association between two measured quantities. Small to moderate but statistically significant correlations were revealed between the following activities:

serving as journal reviewer and ARC assessor (τ =.290, p=.007)

serving as journal editor and editorial board member (τ =.295, p=.007)

serving as journal editor and ARC Panel member (τ =.217, p=.047)

serving as ARC Panel member and ARC assessor (τ =.218, p=.046)

serving for ARC and other external research funding bodies (τ =.363, p=.001)

Professors were most involved in all activities listed in Table 5.3, and senior lecturers the least involved except for serving on editorial boards, where a higher proportion of senior lecturers served than associate professors. Table 5.3 also shows the Chi-square values and probability levels.

Table 5.3 Relationship between Academic Title and Research Assessment Activities

Activity	'Yes' Percentage			Chi-Square	
	Professor	Associate Professor	Senior Lecturer	χ^2	*p*
Have you served as editor of peer reviewed journals?	48.1%	35.7%	31.0%	1.84	.399
Have you served on any editorial boards?	85.2%	46.4%	69.0%	9.40	**.009**
Have you served on the ARC College of Experts?	14.8%	3.6%	0.0%	5.91	.052
Have you served as an ARC assessor?	85.2%	67.9%	20.7%	25.72	**<.001**
Have you served on the assessment panel of any external research funding body other than ARC?	51.9%	57.1%	3.4%	21.46	**<.001**
Have you served on any institutional research committee?	92.6%	75.0%	55.2%	10.16	**.006**

In addition, respondents' research activeness and level of success was also examined, in terms of applying for grant and publishing in peer-reviewed journals between 2001 and 2005 (Table 5.4). The success rates (for example, publications against submissions) reported by individual respondents were high for most. In 2001-2005, 57 (68%) of the 84 respondents had won at least one grant from the ARC and/or the Australian National Health and Medical Research Council (NHMRC); 37 (44%) won competitive grants from other national funding bodies; 40 (48%) won institutional research grants; and 82 (98%) had published in peer-reviewed journals.

Table 5.4 Respondents' Research Activeness and Level of Success

Activity	N of Respondents	Min	Max	Mean
External grants ARC & NHMRC: applied	64	1	35	5.47
External grants ARC & NHMRC: succeeded	57	1	16	3.96
External competitive grants other: applied	44	1	12	3.11
External competitive grants other: succeeded	37	1	10	2.76
Internal competitive grants: applied	40	1	15	3.40
Internal competitive grants: number succeeded	40	1	11	2.85
Manuscripts to peer-reviewed journals: sent	82	3	60	17.96
Manuscripts to peer-reviewed journals: published	82	3	58	16.77

The literature shows that journal acceptance rates tend to be higher in 'hard' sciences than 'soft' sciences. A comparison between education and physics and chemistry was conducted and the results are presented in Tables 5.5 and 5.6. As Table 5.6 shows, there was a statistically significant difference in the acceptance rate of publication between the two fields (p=.018). This suggests that for experienced researchers at senior level, the disciplinary impact on publication success rate persists for this group.

Table 5.5 Respondents' Research Activeness: Comparison between disciplines

Activity	EDUC				P&C			
	N	Min	Max	Mean	N	Min	Max	Mean
External grants ARC & NHMRC: applied	29	1	15	3.28	35	2	35	7.29
External grants ARC & NHMRC: succeeded	22	1	10	2.14	35	1	16	5.11
External competitive grants other: applied	20	1	12	3.25	24	1	6	3.00
External competitive grants other: succeeded	16	1	10	3.56	21	1	5	2.14
Internal competitive grants: applied	17	1	11	2.88	23	1	15	3.78
Internal competitive grants: number succeeded	17	1	10	2.76	23	1	11	2.91
Manuscripts to peer-reviewed journals: sent	47	3	26	7.94	35	12	60	31.43
Manuscripts to peer-reviewed journals: published	47	3	19	6.89	35	11	58	30.03

Table 5.6 Respondents' Level of Success: Comparison between Disciplines

Successful Activity	Total	EDUC	P&C	Comparison	
	Success rate			Chi-square	
ARC & NHMRC grants	71.9%	66.7%	75.1%	$\chi^2(N=57)=26.76$	**p=.013**
Other national grants	78.8%	94.8%	66.7%	$\chi^2(N=37)=9.02$	**p=.029**
Internal grants	86.6%	96.5%	79.3%	$\chi^2(N=40)=14.25$	**p=.014**
Publications	91.6%	89.0%	95.1%	$\chi^2(N=82)=31.45$	**p=.018**

By contrast, respondents from physics and chemistry reported higher success rates in winning grants from the ARC and NHMRC (75.1% compared to 66.7%, p=.013), while those from education obtained significantly more grants from other national funding bodies (94.8% compared to 66.7%, p=.029) and their own institution (96.5% compared to 79.3%, p=.014). This result recognised major differences in the major sources of research funding between the two disciplines.

5.3 Reviewer Perceptions of Journal and ARC Grant Peer Review

Respondents were invited to express their perceptions on journal and ARC grant peer review by rating nine items on a six-point Likert scale, ranging from '1: Not at all' to '6: Very substantially', with slight variations in the wording of the scales for some items (see Appendix 12). The responses were calculated for means and standard deviations. The results are presented in Table 5.7 for journal peer review and Table 5.9 for grant peer review. The rating means for all the items were above 3.5 (neither positive nor negative), which indicated a general positive attitude among the respondents.

Table 5.7 Respondent Perceptions on Journal Peer Review: Means

Item	N	Mean	S.D.
What is your impression of the effectiveness of peer review for journal publication in your field?	82	4.87	.966
To what extent should the practice of peer review aim to improve the quality of a manuscript for journal publication?	84	5.35	.871
To what extent can the practice of peer review improve the quality of a manuscript for journal publication?	84	4.99	1.012
How often do journal editors follow reviewers' recommendations?	83	4.98	.517
Do you bring different expectations to journal and grant peer review?	61	4.96	1.190

In addition, as initially hypothesized, the respondents' view would be influenced by their discipline and experience of reviewing. These two factors, along with the other items (see Tables 5.1 and 5.2), were treated as independent variables and their impact on with the reviewers' perceptions was tested. Six factors – age, gender, academic title, discipline, the total number of papers or applications reviewed, and being an ARC assessor or not – show a statistically significant relationship with the rating.

Tables 5.8 and 5.10 present the outcomes of statistical tests of the items about journal and grant peer review respectively. Statistically significant relationships are highlighted with bold font.

Table 5.8 Respondent Perceptions on Journal Peer Review: Comparison

Item	Age	Title	N of papers reviewed	Gender	Discipline	ARC assessor or not
	One-way ANOVA		Correlation Kendall's τ	Comparing means t-test (2-tailed)		
What is your impression of the effectiveness of journal peer review?	*F*=1.115 *p*=.359	***F*=5.052 *p*=.009**	***τ*=.231 *p*=.011**	*t*=-.087 *p*=.931	*t*=-1.804 *p*=.075	***t*=3.031 *p*=.004**
To what extent should journal peer review aim to improve manuscript quality?	***F*=2.337 *p*=.050**	*F*=.220 *p*=.803	***τ*=.203 *p*=.024**	***t*=-2.971 *p*=.004**	*t*=.527 *p*=.600	*t*=.612 *p*=.542
To what extent can journal peer review improve manuscript quality?	*F*=1.147 *p*=.343	***F*=7.840 *p*=.001**	***τ*=.216 *p*=.015**	*t*=-1.742 *p*=.085	*t*=-1.188 *p*=.238	***t*=3.107 *p*=.003**
How often do editors follow reviewers' recommendations?	***F*=2.683 *p*=.027**	*F*=.307 *p*=.736	***τ*=.215 *p*=.022**	*t*=.196 *p*=.845	*t*=-1.669 *p*=.099	*t*=.926 *p*=.357

The respondents' impression of the effectiveness of journal peer review was highly positive (mean=4.87, Table 5.7). Table 5.8 reveals that the impression was positively related to the total number of papers reviewed (p=.011), the seniority of the academic (p=.009), and whether the respondent had served as an ARC assessor (p=.004). The relationships were all statistically significant. There was no clear relationship between respondents' age or discipline and their perceptions in this regard.

The respondents rated very highly on 'peer review should aim to improve manuscript quality' (mean=5.35). Females displayed a higher expectation than males (5.76 compared to 5.17, p=.004). The expectation of the 45-49 year-old group was marginally higher than those younger than 40 or older than 55 (p=.050). The expectation was positively related to the total number of papers reviewed (p=.024), suggesting that reviewers' positive intention of improving quality intensified rather than faded, with having reviewed more papers. There was no statistically significant disciplinary difference (mean=5.39 compared to 5.29; p=.600).

While the respondents generally thought peer review should aim to improve the quality of manuscripts, they rated 'the extent to which peer review can do so' much lower (mean=4.99, compared to 5.35 for 'aim to improve'; paired-samples t=3.532, p=.001). The difference suggested that, although the respondents generally expected peer review to improve quality, they were much less certain of its capability to

achieve this. There was no difference between age (p=.343), gender (p=.085) or discipline groups (p=.238).

The only between-group statistically significant patterns identified for 'peer review can improve manuscript quality' were related to academic title and being ARC assessors or not. Associate professors demonstrated a more positive attitude than professors and senior lecturers (mean=5.50, 4.96, and 4.52, p=.001), and ARC assessors were more positive than the others (mean=5.27 compared to 4.61, p=.003).

Whether the respondents felt editors followed their recommendations was positively related to their age (p=.027) and the total number of papers they reviewed (p=.022). Physics and chemistry reviewers rated this item higher than education reviewers, but the difference was not statistically significant.

Table 5.9 Respondent Perceptions on Grant Peer Review: Means

Item	N	Mean	S.D.
What is your impression of the effectiveness of peer review for ARC grant allocation in your field?	64	3.92	1.616
To what extent should the practice of peer review aim to improve the quality of an application for an ARC grant?	73	4.52	1.617
To what extent can the practice of peer review improve the quality of an application for an ARC grant?	72	3.88	1.472
How closely do ARC decisions reflect reviewers' recommendations?	55	4.05	1.367

Table 5.10 Respondent Perceptions on Grant Peer Review: Comparison

Item	Age	Title	N of applications assessed	Gender	Discipline	ARC assessor or not
	One-way ANOVA		Correlation Kendall's τ	Comparing means t-test (2-tailed)		
What is your impression of the effectiveness of ARC grant peer review?	**F=3.887 p=.004**	**F=8.318 p=.001**	**τ=.338 p=.009**	t=-.017 p=.986	t=-1.483 p=.143	**t=2.648 p=.010**
To what extent should peer review aim to improve the quality of ARC application?	F=1.205 p=.316	**F=7.150 p=.001**	τ=-.126 p=.341	**t=-2.049 p=.044**	**t=5.132 p<.001**	**t=-4.102 p<.001**
To what extent can peer review improve the quality of ARC grant application?	**F=2.571 p=.035**	**F=3.636 p=.032**	τ=.056 p=.652	**t=-4.486 p<.001**	**t=2.975 p=.004**	t=-1.367 p=.176
How closely do ARC decisions reflect reviewers' recommendations?	**F=3.534 p=.008**	**F=7.803 p=.001**	**τ=.491 p=.000**	t=-1.535 p=.131	t=-1.064 p=.292	**t=2.561 p=.013**

While in general the respondents have given fairly high ratings to all the four items about journal peer review (means=4.87, 5.35, 4.99, and 4.98), the ratings of the four items related to ARC grant peer review were much lower (means=3.92, 4.52, 3.88, and 4.05; see Table 5.9). The differences were all statistically significant at .001 level.

The mean score of the effectiveness of ARC grant peer review was 3.92 (Table 5.9). A statistically significant higher rating was given by those who were ARC assessors (p=.010), while the 50-54 age group (mean=2.43, between group p=.004) and senior lecturers (mean=2.91, p=.001) rated this item considerably lower than the other groups. No difference by discipline was identified (p=.143) (see Table 5.10).

With regard to whether peer review should, and can, improve the quality of grant projects, education respondents demonstrated a significantly higher expectation than physics and chemistry respondents (Table 5.10). In general, the respondents' expectations of grant peer review in this regard were much lower than that for journal peer review. Table 5.11 presents an overall comparison, and Tables 5.12 and 5.13 provide a breakdown by discipline. Compared to those from physics and chemistry, education respondents showed a clearer expectation for peer review to improve the quality of applications, and believe in its ability to achieve so. While education respondents' ratings between journal and grant were not different (Table 5.12), those from physics and chemistry rated grant peer review for these two aspects considerably lower (Table 5.13).

In addition, the respondents' belief in peer review's ability to improve application quality was much lower than their expectation of its 'aiming to improve' (paired-samples t=3.646, p=.001), not unlike the result for journal peer review.

Respondents who were ARC assessors appeared to be less supportive of the idea of using peer review to improve the quality of applications (mean=4.00, compared to non-assessors 5.31, p<.001). The same pattern was found between males and females ('aim to improve': 4.33 compared to 5.27, p=.044; 'can improve': 3.49 compared to 5.12, p<.001); and between associate professors and the other two groups ('aim to improve': 3.57 for associate professors compared to 4.80 for professors and 5.12 for senior lecturers, p=.001; 'can improve': 3.32 compared to 4.41 and 3.78, p=.032).

Table 5.11 Respondent Expectations of Journal & Grant Peer Review: Overall

	Journal peer review	**Grant peer review**	**Comparison** (Paired samples t-test)	
Statistics	Mean	Mean	*t*	*p*
Should aim to improve quality	5.34	4.52	4.165	**<.001**
Can improve quality	4.92	3.88	5.034	**<.001**

Table 5.12 Respondent Expectations of Journal & Grant Peer Review: Education

	Journal peer review	**Grant peer review**	**Comparison** (Paired samples t-test)	
Statistics	mean	mean	*t*	*p*
Should aim to improve quality	5.45	5.24	1.198	.238
Can improve quality	4.70	4.35	1.379	.176

Table 5.13 Respondents' Expectations of Journal & Grant Peer Review: Physics & Chemistry

	Journal peer review	**Grant peer review**	**Comparison** (Paired samples t-test)	
Statistics	mean	mean	*t*	*p*
Should aim to improve quality	5.19	3.55	4.718	**<.001**
Can improve quality	5.14	3.37	6.231	**<.001**

Whether the respondents felt ARC decisions reflected their recommendations was positively related to their academic seniority (title) (p=.001) and the number of applications they assessed (p<.001). The 50-54 age group rated it considerably lower than others (mean=2.82, p=.008). It should be noted that this group had also rated the effectiveness of ARC peer review significantly lower than the other age groups.

Sixty-one respondents had reviewed for both journals and the ARC. In general, they brought different expectations to reviewing in the two contexts (mean=4.96 out of 6). Certain groups differentiated their expectations to a greater extent than others, including females (t=-2.249, p=.037), associate professors (F=3.131, p=.050), those newer to the job of serving as ARC assessors (τ =-.336, p=.013), and those who were younger than 40 (F=2.912, p=.022). No disciplinary difference was identified.

The respondents were invited to judge which one, grant or journal reviewing, was the more important activity for them. Among the 58 who offered an answer, 35 thought journal peer review was more important, seven thought grant peer review was more important, and 16 felt the two were equal in importance.

The respondents were also invited to offer an evaluation of their overall experience with journal and grant peer review. The mean scores were 4.61 for journal review and 3.90 for grant review. The difference was statistically significant (t=-4.352, p<.001). Among the groups of comparison, non-ARC assessors and senior lecturers displayed the lowest satisfaction with both experiences; the differences being both statistically significant. No difference was identified between the other groups.

Table 5.14 Respondents' Level of Success & Overall Satisfaction with Peer Review

Successful Activity	Satisfaction with journal peer review		Satisfaction with grant peer review	
ARC & NHMRC grants	r=.370	**p=.006**	r =.249	p =.087
Other national grants	r =.055	p =.746	r =-.223	p =.211
Internal grants	r =-.159	p =.327	r =-.216	p =.212
Publications	r =.391	**p =.000**	r =.373	**p=.003**

Pearson correlation coefficients were calculated to test for the relationship between the respondents' level of success and their satisfaction with peer review. The results are presented (Table 5.14). Significant relationships were identified between the level of satisfaction with journal peer review and the success rates in gaining ARC and NHMRC grants and success in getting published , and between satisfaction with grant peer review and publication success.

5.4 Motivation to Participate as Reviewers in Peer Review

Rigorous peer review takes considerable time and effort. The survey results show that none of the respondents had been paid by any journal for reviewing. The survey included some items to find out why they were willing to do the job, and whether and why they declined to do it at times. A comparison was also conducted with their motivations for contributing to ARC grant peer review.

A majority of the respondents (64.3%) had refused to review, either some papers or for some journals and they gave a total of 100 reasons for doing so (Table 5.15). The most cited reasons for declining to review some papers included: the paper was outside their field of expertise, lack of time, and conflict of interest because they knew the authors. The most cited reasons for declining to review for a journal included: lack of expertise, lack of time, a perception that the journal was of poor quality, and their experience of being treated badly by the journal before. There was no clear disciplinary difference identified in citing the reasons.

Table 5.15 Reasons for Declining to Review Some Papers or for Some Journals

Reason	Frequency (%)		
	For paper	**For journal**	**Total**
Lack of expertise	37 (57.8%)	14 (38.9%)	51 (51%)
Lack of time	23 (35.9%)	10 (27.8%)	33 (33%)
Other	4 (6.3%)	12 (33.3%)	16 (16%)
Total	64 (100%)	36 (100%)	100 (100%)

With regard to reasons that motivated them to review for journals, 95 reasons were given, falling into five groups (Table 5.16) where paraphrasing and quotations are used, the ID of the respondent is included in square brackets. The respondents perceived their contribution firstly as a professional obligation and viewed it 'an expectation of academic job' [1189]. Some viewed peer review as a good opportunity for professional development, in the sense of 'getting a better feeling for what papers are accepted' [2138], 'keeping abreast of current developments in the field' [2184], as well as 'gaining something to put on CV' [1159]. Some reviewers were motivated to contribute their time and expertise as a form of 'collegial support' – assisting the editor [1150] and the author [1244], and sharing the load with other academics [1260].

As Table 5.16 shows, a higher proportion of education reviewers noted collegial support as a motivation. Six respondents noted maintaining the quality of journal publication as a motivation for reviewing.

Table 5.16 Motivation to Participate as Reviewers in Peer Review: Frequency (%)

Reason	Education	Phys&Chem	Total
Professional responsibility and obligation	19 (35.8%)	18 (42.9%)	37 (39.0%)
Personal professional development	18 (34.0%)	13 (31.0%)	31 (32.6%)
Collegial support	9 (17.0%)	3 (7.1%)	12 (12.6%)
Being interested or expert in the topic	5 (9.4%)	4 (9.5%)	9 (9.5%)
Maintaining the quality of publications	2 (3.8%)	4 (9.5%)	6 (6.3%)
Total	53 (100.0%)	42 (100.0%)	95 (100.0%)

Table 5.17 Reasons for Agreeing to Review for the ARC: Frequency (%)

Reason	Education	Phys&Chem	Total
Professional responsibility & obligation	12 (41.4%)	22 (62.9%)	34 (53.1%)
Personal professional development	7 (24.1%)	7 (20.0%)	14 (21.9%)
Collegial support	6 (20.7%)	0	6 (9.4%)
Being interested or expert in the topic	2 (6.9%)	3 (8.6%)	5 (7.8%)
Ensuring high quality grants get funded	2 (6.9%)	3 (8.6%)	5 (7.8%)
Total	29 (100.0%)	35 (100.0%)	64 (100.0%)

Those who had served as ARC panel members or assessors also provided reasons for contributing to grant peer review; 64 reasons were given, also falling into five groups (Table 5.17). A few respondents indicated they contributed to grant peer review because it was 'a condition of grant awards' [3165]. None of the physics and chemistry respondents cited collegial support as a motivation. Discipline interest and a concern for ensuing that high quality grants are funded were rarely mentioned.

Moving to journal policies, 85 percent of the journals that the respondents reviewed for almost always provided them with guidelines (Table 5.18). Their motivation to follow the guidelines was fairly high (overall rating score=3.49 out of 4); 93.8 percent of the respondents noted that they almost always followed the guidelines provided (Table 5.19).

Table 5.18 Frequency of Journals Providing Guidelines to Reviewers

Scale	Some	Most	All	N/A	Total
Frequency	10	29	39	3	81
%	12.3	35.8	48.1	3.7	100

Table 5.19 Frequency of Respondents Following Journal Guidelines

Scale	Always	Mostly	Sometimes	Never	Total
Frequency	47	29	3	2	81
%	58.0	35.8	3.7	2.5	100.0

The relationship between the extent to which the reviewers followed guidelines and their perceptions of the usefulness of the guidelines in helping them generate better reviews was significantly positive (r=.342, p=.001). The rating score of the usefulness of the guidelines (4.44 out of 6) indicated a moderate positive perception. The perception was related positively to the experience in reviewing, including the total years of serving as a reviewer (Kendall's τ =.212, p=.016) and the number of papers reviewed (τ =.272, p=.002). By contrast, respondents who were editors (t=-2.000, p=.049) and ARC panel members (t=-2.080, p=.041) rated the usefulness of the guidelines significantly lower than those who were not. There was no clear difference by discipline or sex.

5.5 Additional Comments on Peer Review

Towards the end of the survey, the respondents were invited to make additional comments about peer review, and 28 shared their thoughts. Some respondents used the opportunity to confirm peer review as the key to maintaining quality [1150] and ensuring quality and integrity in publication and grant writing [2184], while some others identified concerns about peer review.

The concerns included that: subjectivity [1104] was inevitable in peer judgment; the process sometimes took an unnecessarily long time [1159] and substantially slowed the publication process [1194]; some reviewers were very irresponsible [2166] and

some reviews were disappointingly brief and surprisingly unhelpful [1208]. One respondent stated that:

> The peer review process works well when reviewers are concerned to promote scholarship and knowledge in an area of study. It does not work well when it is used by reviewers to exclude promotion of alternative views other than their own or when reviewers are not sensitive to the feelings of authors and write inappropriate and sometimes offensive reviews. [1240]

Another respondent described his experience as an author, feeling disadvantaged by the anonymous review system:

> On a number of occasions, I have submitted a paper to a journal, had it rejected, submitted it again to a different journal and had it accepted. This is not explained by the rejecting journals having extremely high standards, but by unprofessional refereeing. The anonymous review… has major flaws in my field, which is badly affected by ideological wars. The current practice of anonymous reviewers and non-anonymous authors encourages reviewers to hide behind their anonymity. [1189]

Several respondents expressed concern about the lack of preparation for new reviewers, indicating academics should have the necessary skills to peer review, but these skills are rarely (if ever) discussed [1226]. There was also criticism of publishers in that they made use of academics without paying them yet charged academics and libraries extortionate fees to access works whose quality was ensured by academics [1159].

Many respondents noted the increasing difficulty to find time to review promptly or thoroughly. Some emphasised they had always tried to give timely and useful feedback because they understood the review process can take too long [1150]. Other reviewers highlighted the heavy workload and time pressure they were facing and appealed for greater recognition of their contribution by employing institutions as an important aspect of academics' work [1210].

5.6 Conclusions and Discussion

The initial survey collected the opinions of a group of experienced and successful academics. A sizable proportion of them were from leading universities in Australia, had the experiences of serving as journal editors, editorial board members, or ARC assessors, and all had reviewed for journals for at least four years.

The respondents' involvement in research assessment activities reinforced their level of success when going through such activities themselves (for example, reviewing compared to being reviewed). They had been active in terms of applying for competitive grants and submitting papers for publication, and their success rate was generally very high. They provided invaluable insights into the peer review system of two distinct fields.

Disciplinary comparison was a core focus in the analysis of the survey data. However, no difference was identified in regard to either the respondents' perceptions or expectations of journal peer review. The only statistically significant difference that was identified was in the expectation for grant peer review to improve the quality of grant projects, where education respondents provided higher ratings.

However, the survey results also showed that respondents who were ARC assessors were significantly less supportive than non-assessors of the idea of using peer review to improve the quality of applications. In the survey cohort, there were a higher proportion of physics and chemistry respondents than education ones who had served as ARC assessors. When these two factors were considered, it would appear that the disciplinary difference was more related to the characteristics of the participants. Other than this difference, the respondents from the two broad fields demonstrated very similar perceptions with regard to all the aspects being studied.

Other factors than discipline, such as gender, academic seniority and research experience, showed stronger influences in the differences identified in the data. For example, more experienced reviewers and ARC assessors, in terms of the total number of papers or applications reviewed, gave considerably higher ratings on the effectiveness of peer review, their satisfaction with the system, the usefulness of

'guidelines for reviewers' in generating better reviews (referring to journal reviewers only). They were also more certain in the belief that the editorial or ARC decisions reflected their recommendations.

Gender was another influential factor. Compared to males, females were more supportive of the idea that peer review should and could improve the quality of manuscripts and applications, and they were more inclined to bring different expectations to reviewing for journals and grants.

The respondents' level of success was also a factor of influence. The more successful they were in research assessment, especially in terms of winning competitive grants and publishing in peer-reviewed journals, the more satisfied they appeared to be with the current system, irrespective of their discipline, gender, or academic title.

A marked difference was identified between journal and grant peer review. Journal peer review was rated statistically significantly higher than grant peer review for all compatible items, including: respondents' expectation regarding whether peer review should and could improve the quality of manuscripts, their perception of the effectiveness of the two systems, the extent to which their recommendations were reflected by the final decision, and their overall satisfaction with the two systems.

The survey also afforded insights into the respondents' motivations to contribute to the process. Journal reviewers viewed their contribution mainly as a professional obligation and an opportunity for self development, while ARC assessors noted that assessing others' applications was a condition of their success in gaining grants. Considerations connected with providing collegial support or improving quality were not strongly cited as motivations for grant peer review.

Most of the respondents had at least once refused to review some papers or for some journals. While 'lack of time' was a major and obvious reason for doing so, the lack of expertise was the most cited reason. This indicated that many reviewers had the experience of receiving papers on a topic outside their field of expertise.

This seemed to be a common problem in peer review. Authors' responses reported in several studies suggested that the editor might not necessarily know who were competent reviewers in a particular area (Bedeian, 2003; Bradley, 1981; Epstein, 1990; Nicholas, et al., 2006; Thyer & Myers, 2003), which, according to Callaham and Tercier (2007b), was a 'crucial limitation in the peer review process' (p. 38).

To conclude, the initial survey expanded the understanding of journal peer review. The issues it covered, including reviewers' perceptions of the functions of peer review, the ways in which they approached reviewing, and any change they thought should take place in peer review, were explored in further phases, including the interviews, the follow up survey, and the analysis of reviews.

CHAPTER 6: INTERVIEW PARTICIPANTS' COMMENTS ON THE PRACTICE OF PEER REVIEW

6.0 Introduction

In the field of peer review interview-based studies are few and do not focus on journal reviewers specifically; some focus on editors (Weller, 1990a), grant panelists (Guetzkow et al., 2004) and authors (Misakian & Bero, 1998), and others focus on a general group of academics (Calnan et al., 2006; Flowerdew, 2001; Sutherland et al., 1993). Two studies used interviews to explore criteria for evaluating quality: one focused on the criteria used to judge clinical research questions and hypotheses (Sutherland et al., 1993) and the other examined how fellowship panelists in humanities, history, and social sciences defined 'originality' (Guetzkow et al., 2004).

The complexities associated with peer review, including those arising from the different perspectives of the actors were highlighted in the previous chapters, and are touched on in the survey findings in Chapter 5. In this study the reviewers who answered the initial survey were asked if they were prepared to be involved in a semi-structured interview to explore their perceptions of peer review more fully (see the interview protocol in Appendix 13). This chapter focuses on the first group of six questions that sought to elicit their general expectations of peer review, how they developed their skills as reviewers, how they believed their reviews were used, and what developments they had perceived or wanted to see in practice. The questions are:

- What do you think is the major contribution of peer review to your field?
- What is your personal expectation of peer review?
- Have you noticed any trends in peer review processes over time?
- Is there anything you would change in the processes if you could?
- How did you learn how to undertake peer review?
- Are you aware how closely editors follow your advice?

This chapter is organised in the sequence of each of the questions, and responses are further grouped into categories where pertinent. Interesting points raised in the responses to these questions were probed further as themes in a follow-up survey and are reported in the final section (6.10).

6.1 Participants in the Interview

A total of 44 academics were interviewed (Table 6.1). There was no statistically significant difference between those who responded and who did not by discipline, suggesting that the current cohort, although small, remained a proportional sample of the population of Australian senior academics in education, physics and chemistry.

Table 6.1 Interview Participants from Education, Physics and Chemistry

Discipline	Professor		A/Professor		Senior Lecturer		Total	
	male	female	male	female	male	Female	male	female
Education	6	1	6	1	2	9	14	11
Physics & Chemistry	6	1	4	0	8	0	18	1

The responses were de-identified during transcription with four-digit ID codes. The first digit from the left denotes discipline, 1 for education, 2 for chemistry and 3 for physics; the second digit denotes sex, 1 for male and 2 for female, and the third and fourth digits are serial numbers assigned to each individual. The IDs are used in connection with quotations throughout the chapter and direct quotations are slightly edited and modified in minor ways to protect identity.

6.2 The Contribution of Journal Peer Review

Most of the participating reviewers cited two or more types of contribution of journal peer review. In total six types were identified, falling broadly into two groups – 'contribution to scholarly communication' (see categories 1-3 in Table 6.2) and 'contribution to the individual' (categories 4-6 in Table 6.2). In the latter this refers to reviewers and/or authors. The table provides illustrative quotation reflecting each category. Categories closely reflect, or are the exact, words used by informants and selected as coding categories (a process facilitated through the use of *QSR NVivo*).

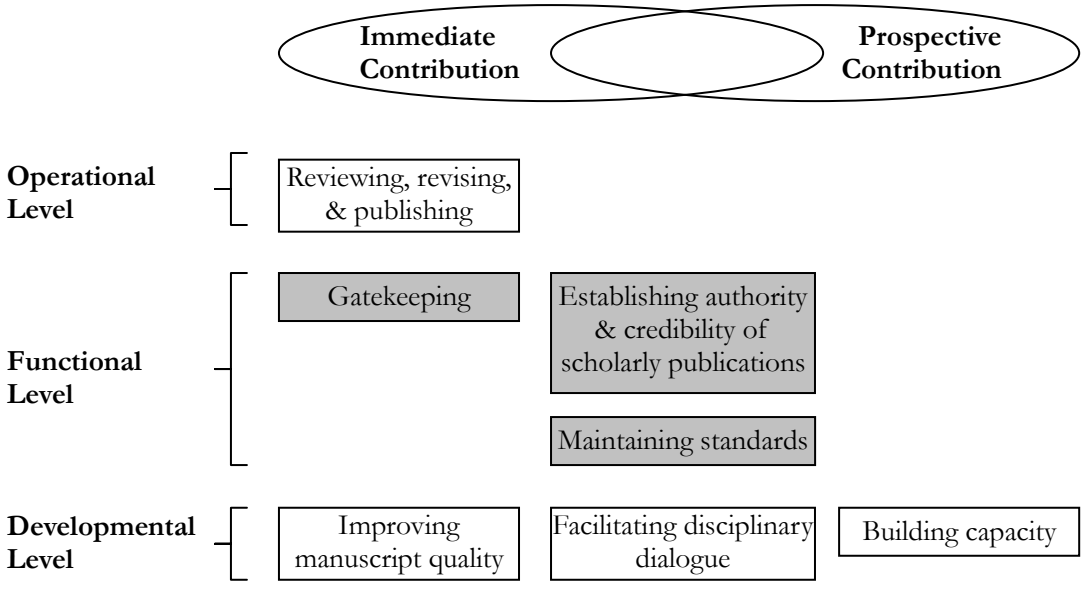
Table 6.2 Contribution of Peer Review to a Field

Category	Example
1. Maintaining standards	It ensures there is…quality associated with the research in your field. [1133]
	I certainly believe that peer review contributes to the maintenance of quality of publication. [1224]
	It is to maintain the standards of the journals. [3103]
	It is to establish a minimum standard of integrity and quality for scientific publications. [2189]
2. Gatekeeping	It makes sure that work that is substandard doesn't get through to publication. [3181]
	to keep the work out that is not up to publishable quality [2129]
	It doesn't mean it's always right, but it does mean someone has thought about it and the rubbish has been filtered out to some degree. [3145]
3. Establishing credibility & authority	It establishes some sort of credibility for the field you're working in, in terms of the research output for that field. [1194]
	Through the peer review process, I believe that only good quality research in my discipline is published. [1224]
	It ensures that, within reasonable limits, those persons performing the research do so using methodologies that are acceptable within the particular area of research that's being used; and that the people who are engaged there are recognized as being qualified by the research area. [1151]
4. Building capacity	Peer review provides primarily a process supporting colleagues, particularly early career colleagues through evaluating and providing feedback to their work… You are helping them to develop their skills and competencies necessary to advance their career. [1126]
	It also ensures that less experienced authors get some advice about their publications and the appropriateness of their thinking. [1146]
	It gives you good feedback of how your publications are being assessed in that particular area and also whether your research is worthy of doing. So, you get an expert's feel or an expert's opinion on your research. [2138]
	That's about keeping up to date for everybody; not just those who are being reviewed, but also for those like yourself and it's probably one of the major things we get out of doing peer review and being reviewed. [1202]
	There's almost a sense of a professional development role, too. If people are involved in the peer review process, they probably develop a sense of an understanding of the range of the types of research that are going on. [1208]
	It contributes to the professional development of all the researchers involved in the processes of peer review… It's all about capacity building. [1255]

Category	Example
5. Improving manuscript quality	It's a good mechanism for fair and thorough feedback that can advance the quality of publications. [1210]
	I see it as an iteration the aim of which is to make the paper as good as possible. [3183]
	It adds strength to the work that is to be published. [1224]
6. Facilitating disciplinary dialogue	Essentially it allows people with the same level of expertise in largely the same field of expertise to comment on your work. [1154]
	You give a report, it goes to the editor, they send it to the author, they make a reply, it goes back to you, you have a conversation and the aim of the conversation is to make sure that what is published is correct. [3183]

In general, the emphasis was weighted toward contribution to scholarly communication over contribution to individuals, and here 'maintaining standards in publication' was most cited, followed by 'gatekeeping' and 'establishing credibility and authority of published work'.

Figure 6.1 A Function Schematic of Peer Review



Education reviewers were more likely to identify contribution to the individual than physics and chemistry reviewers. Three education reviewers cited 'improving quality of manuscripts' as the major contribution; ten reviewers from education but only one from physics cited 'building capacity'. This difference suggests that education

academics value peer review as an opportunity to 'educate' authors and 'develop' those involved. While some from physics and chemistry also signified the educative role (for example, [2138]), it was not typical.

In the discussion of contribution, a conception of 'What is peer review?' gradually took form. At the operational level, peer review is an 'iteration process' [3183] involving the back and forth dialogue between author, reviewer, and editor. At the functional level, it is 'the baseline measure of quality' [1185] which 'develops a sense of standards across the community of scholars' [1150] and 'ensures and raises the quality of publications' [1210], so that 'the knowledge to be relied on' is 'authentic, valid, and reliable' [1249]. And at the developmental level, it has the function of facilitating dialogue and building capacity among researchers in a field who are often scattered geographically, and improving individuals' work and research.

The contributions can also be further classified into immediate contributions to journal publication and prospective contributions to scholarly communication and the wider research community. Figure 6.1 presents a schematic of the conception.

The nature and functions of peer review as described above are closely interrelated with the holistic concept of 'stewardship' of disciplines. According to Walker et al. (2008, p. 12), stewards are 'aware of the shoulders on which they stand and must judge which ideas are worth keeping and which have outlived their usefulness' and think about 'the continuing health of the discipline and how to preserve the best of the past for those who will follow'. Responsible, qualified reviewers are important stewards, who 'are concerned with how to foster renewal and creativity' and consider 'how to prepare and initiate the next generation of stewards'. In this sense, peer review empowers its participants. As one respondent indicated,

> It skills all of us; it's capacity building for all of us in responding really carefully to each other's work; it contributes to the professional development of all the researchers involved in the process of peer review. [1255]

When talking about the contributions of peer review, the tone of most reviewers was confirmative. They often noted the 'very valuable role' [1169] peer review played in

their fields, viewed it as 'absolutely essential' [2107] and 'absolutely crucial' [1221]. It conveyed their faith in peer review. This sense of faith has been expressed so strongly throughout the interviews, and generally in the literature (for example, the editorials), that this topic is revisited later in the chapter.

6.3 Expectations of Journal Peer Review

Reviewer responses about expectations of peer review fell into three major areas, including the qualifications, competency and professionalism of reviewers, the quality of reviews and their usefulness for improving manuscript quality.

Respondents expected to provide authors with a high quality peer review process, in which they were treated fairly, received constructive comments, only the merit of their work not other things was evaluated, and their papers were improved as a result. The expectations are aligned to the provision of adequate stewardship, in that reviewers maintain and advance their field and mentor new generations of stewards.

They expected reviewers to be 'honest' [2138] and to have 'reasonable familiarity' with the research area [2153], and to act as 'a critical friend' [1126] [1154], offer 'forthright and constructive criticism' [1146], 'engage with the ideas' and 'extend or refine those ideas without resorting to attacking you [the author] because of a particular theory or method or idea' [1154]. The word 'constructive' re-occurred in the responses. They also expected reviewers to read a paper 'really carefully' [1255] and to 'really understand it' [2169]; yet one admitted that, 'not everybody thinks so and some think this ought to be the author's responsibility. There is too much work involved just reading and checking every detail' [3110].

Clear and specific comments were desired. As one noted, 'It's a total waste of time if you get information back from a journal which simply says, 'Your paper has not been accepted' or 'Your paper has been unsuccessful' and that's all' [1147]. And as another noted, 'if the referees have problems with [the paper], I would like them to elaborate on what they think the shortcomings are, why they don't think it is "cutting edge" research, and what they think should be cutting edge research.' [1163]

Respondents expected reviewers to be fair and 'be reasonably flexible and tolerant of different approaches and points of view' [1127] and 'make judgment entirely upon the basis of the research and the documentation of that research rather than being 'tainted by any sort of personal interaction' [1218]. One noted the difficulty of avoiding this issue in Australia 'where we are relatively small and where people know each other, unless the review is blind, which is often difficult to achieve, there can be a highly personal component in it. That could be very negative.' [1185]

Reviewers need to 'not just be identifying weaknesses but to be pointing the author to possible strategies or sources to pursue in terms of addressing those weaknesses' [1210]; 'approach a paper with a view to improving it or with a view to pointing out where it might need further work' [1208]; 'pick up items or parts that are not clear' [2189]; and 'devote some time and effort to improving the work, with suggestions on how to improve the work' [3181].

Some emphasised that, as reviewers were selected for their expertise in a field, they should assist authors by defining the 'cutting-edge' of the field, for example, if 'there has been other research on that particular point which the paper doesn't acknowledge or use, the reviewers have a function there of pointing that out and suggesting where it might be useful to follow up on other research that might affect the argument.' [1150]

When respondents took the standpoint of authors in their comments, they appeared to be more critical about the process: 'I expect peer review to be fairly quick… the process has never been really quick' [2173] and 'I expect that peer review is a supportive process. That is not what normally happens' [2177]. Several respondents from physics and chemistry wanted the process to be quicker. One noted his own experience of a period of several years between submission and publication of his paper [2173]. Another noted: 'there is nothing worse than submitting a paper to a journal and having people sit on it for a long time for various reasons' [3145].

Of the respondents, 19 had the experience of editing journals. They were asked in addition: 'As an editor, what is your expectation of reviewers?' Their responses did

not go beyond what they said previously, such as the expectations of the quality of reviewers and reviews. Many made a statement of this type: '[my expectation as an editor is] exactly the same as what I expect to do as a reviewer' [1244]. As such, these responses were incorporated in the foregoing discussion.

6.4 Changes Sought

Twenty-six of the respondents (59%) indicated they would like to see changes to the current peer review system to improve it and 18 (41%) indicated they would not. As Table 6.3 shows, those who said 'yes' had a little longer experience of reviewing than those who said 'no'; and a slightly higher proportion of education respondents than those from physics and chemistry opted for 'yes'.

Most respondents offered justifications for their responses. Among those who said 'no', nine indicated that they were satisfied with the current system, on the basis that 'The process is pretty straightforward… It is well understood by practitioners in the field and researcher' [1126]. Some others thought the system as a whole worked well and the problems rest with 'individual cases which relates to the behaviour of individual referees' which 'comes down to human nature' [2184].

Table 6.3 Respondents' Yes-or-No Responses to 'Changes to Peer Review'

Discipline	'yes' Respondents		'no' Respondents	
	N (total n=26)	Years of reviewing	N (total n=18)	Years of reviewing
Education	16 (61.5%)	12.7	9 (50%)	11.9
Physics & Chemistry	10 (38.5%)	17.7	9 (50%)	14.7

The 26 respondents who answered 'yes' pointed out a range of areas where change was needed, including: the efficiency of the process, the clarity of journal guidelines, editor leadership, reviewer recruitment, blind review, reviewer comments, and the recognition of reviewers' contribution.

Lack of efficiency was recognised as a major issue of journal peer review. In order to solve it, one suggested the establishment of a 'central reviewing agency' where all

submissions were reviewed and then journals 'bid for papers through this agency', which allowed papers to be 'published in the best journal' without 'losing any time going through this process' [2107]. Another suggested that authors' institution have an internal review process to 'ensure that the paper has been looked at by colleagues before it has been sent out' [3102]. A third suggested editors remind reviewers more often to return their reviews [1169].

'Editorial voice', in the sense of providing either a clear statement of the scope of the journal or specific instructions for reviewers, was perceived as lacking. As one noted, 'journals need to stand for something; they need to have an editorial voice; they need to have a clear idea of the kind of contribution they want to make to a particular field', rather than being 'so broadly based that they will accept papers with almost any kind of research, with almost any kind of methodology, with almost any kind of topic, [and] with almost any kind of voice' [1146]. Some noted that 'sometimes the criteria that we give to reviewers can be a little bit vague' [1224], which resembled 'four broad brush strokes' [1202]. When journals failed to specify the criteria, the reviewers felt it hard to 'establish a profile to provide to the editor' [1159] or review in a 'consistent' manner [1224].

Some respondents criticised editorial inclination 'to abdicate responsibility and lean on the reviewers in particular' [3103] and urged them to take a stronger leadership by providing timely feedback to reviewers, and exercising their 'responsibility and authority' [3103] to make a decision when reviewers did not agree with each other.

One noted the editors' role in facilitating post-review communication between reviewers and authors, 'I would like the opportunity to engage, through the editor, who could ensure anonymity, with a reviewer, where I feel the author has missed the point I was making or where I have not understood the point the reviewer is making' [1127]. Another noted the necessity for editors to let reviewers know the final decision 'even if they choose not to use' their reviews, and viewed editors who failed to do so as 'rude' [3183]. Again another noted the editor's role in encouraging new researchers by giving them 'feedback so that they can learn from the process', and offered an idea to let authors 'tick a box' if they were early career researchers. [1224]

Concerns about recruiting appropriate reviewers were also evident in the responses. Some expressed a 'trust' of editors in this and felt 'editors are really careful and do their job well' [1260], but more noted difficulties in achieving a balance between expertise in an established area and receptivity to new or alternative ideas, between the expertise of multiple reviewers, and in 'finding a combination of reviewers who are on the one hand offering a range of assessments, but on the other hand, not so disparate as to lead to radical disagreement.' [1150]

As one indicated, often a paper is sent to 'some one who works in a very different theory or theoretical framework with which the paper is written', and then the author would receive reviews that 'attack or criticize the theory or the theoretical methodology rather than the article itself.' [1208] A strategy to recruit reviewers with balance of expertise was offered by an education respondent [1151]; the general idea was to employ 'someone who is in an entirely different research area' from one of the reviewers to 'allow for greater flexibility and… for the recognition of change in the whole paradigm movement.'

Some issues with the practice of anonymous review were also identified in the responses. One respondent indicated that anonymous review 'creates all kinds of problems [and] double blind is a better system' [2107]. Another associated anonymous review with poor quality reviews, 'that facelessness is a little bit of a license to make comments without justifying them.' [1202] A third felt that extra care needed to be taken to ensure 'true' double-blinding, 'Sometimes the authors and the editors need to give more attention to masking the identity of the author, so that blind reviewing is, indeed, blind.' [1249]

Several respondents noted the need to improve the quality of reviews, but only one offered a strategy to do so: a policy should be 'written into reviewing documents that says: 'If you speak negatively or particularly positively about this aspect of the piece, would you justify the comments?' [1202]

In addition, several respondents raised the issue of inappropriate recognition of reviewers' contribution: 'the process, the time, and the workload don't feature anywhere' [1208]; 'there's no credit for it; there's no space in our workload for it. It

is slightly a secret work.' [1255] One expressed her concern about 'the danger of overloading' good reviewers and losing them due to the lack of recognition and increasing pressure in the sector [1210]. Using reviewers' service for free by 'journals that can make a profit' was viewed as 'unfair' [2107], while more indicated that appropriate recognition rather than payment was what was needed. Some maintained that peer review was a 'professional obligation' [2166] and monetary payment would run counter to the moral of it. It was suggested that 'the process needs to be recognized as part of an academic's job [and] professional academic workload.' [1208] and 'it can be a similar process as when you are feeding back to a student's working process… and it ought to be part of how it skills us up to do that work' [1255].

Responses to this question once again displayed respondents' faith in peer review. Although most of them were critical about the process and indicated a desire for change, none challenged the baseline process, that is, to invite external experts to evaluate and, based on their advice, to let the editor make the final decision.

Another noted problems caused by using only one reviewer per paper in his field (physics). But, when asked whether he wanted to see a change to this, he replied that, 'It takes a long time for everyone to review a paper… and they have trouble finding people to referee papers… The question is 'Why two? Why not five? Why not ten?' you have to stop somewhere' [3183]. The cohabitation between criticism and faith is deeply woven into the academic fabric of peer review.

6.5 Trends Noted

The respondents were asked whether they had noticed any trends or changes in peer review. Spontaneously, 14 from education and nine from physics and chemistry responded 'no', yet all of them had noted at least some minor developments. Three groups of changes were identified, as shown in Table 6.4. Most of the respondents observed an increased pressure to publish in universities which had led to a marked growth in the number of submissions, a decline in work of good quality, and an increasing demand for reviewers. Some noted a tendency for reviewers to be more likely to decline to review and reviews had become shorter and less detailed; recommendations had become harsher, and more papers were rejected.

Table 6.4 Trends in Journal Peer Review

Response group	Education (N=25)	Physics &Chemistry (N=19)	Total (N=44)
Pressure to publish	20 (80.0%)	14 (73.7%)	34 (77.3%)
Electronic handling	3 (12.0%)	6 (31.6%)	9 (20.5%)
Journal approach	6 (24.0%)	1 (5.3%)	7 (15.9%)

One noted the link between the pressure to publish and income generation for universities and that two consequences of the pressure were that academics tended to 'send off very poor quality work, just trying to get a lot of work out', and 'any failure to publish causes a lot of stress' [1154]. It was observed by another that it was unfortunate but under such pressure some authors rush in and are inclined to 'use reviewers to finalize the papers' [3102].

Another consequence of the higher demand on reviewers was seen to be impoverished reviewing. As one noted,

> People simply don't have time, enough time, to properly review, so they just skim through papers. In some cases, just try to find a hole in it and get rid of it instead of trying to do it properly. That is the worst thing that has happened in the last 20 years. [3180]

One reviewer explained changes in his own practice to show how quantity militates against patient and formative feedback:

> When I first got papers ten years ago to review…I would scribble on them extensively to suggest work changes and suggestions to make them considerably better. I don't do that anymore. I don't think that editors want that. They want your opinion of the broad brushstrokes of the paper. They want you to act as editor. I also stopped doing it because of the number of papers I am asked to referee makes it impossible. [3181]

Increasing demand, more substandard papers, and their own need to publish can detract from the inclination to act as a reviewer. Some responses reflected reviewers' struggle between collegial obligation and their individual availability, 'Really you do

face the issue with the time spent reviewing a paper you could be actually writing one yourself, and the pressures are on to be writing not reviewing.' [1210]; 'the pressure is too big; people should be more selective about which papers they accept for review', however, 'the whole system relies on peer review… It's an obligation that we all have to participate in; otherwise the whole system will collapse' [1218].

Another trend identified in the responses is electronic submission and reviewing. Some noticed new efficiencies that this change had brought about, noting that 'nowadays, you get an email asking you to do it; you can access the paper immediately, the return is done online.' [3145] The wide use of emails allows 'more scope for further iteration between the reviewers and the authors as how to improve the paper… there is more communication as the reviewers send more information and the authors responded to that, and then there is further iteration.' [2189]

Not everyone regarded the transition to new technology as a smooth one. The preference of one respondent was hardcopy: '[it] makes it harder for the reviewer when you have to master the journal's electronic system, where you have to waste time getting into it, do the review at your computer, where you are less likely to be able to concentrate' [1244]. Another potential negative effect of electronic reviewing could be 'that the reports you get back… are very brief, sometimes just a few lines' [2215]. But as this respondent indicated, shorter reviews were mainly a result of increasing time pressure on reviewers; the electronic system just worked as a catalyst.

Again another marked trend was that journals had become more 'systematic' [1150] and 'organised' [1185] [3110] in the way they approach peer review. An example provided was 'almost universally, there are checklists and criteria are specified and sometimes there are rating scales associated with them, [and] they are much more precise and structured.' [1150] One respondent associated this with the accelerating demand to publish in peer-reviewed journals which led to 'tightening' and 'lifting the game' of the peer review process for more journals which had 'always been there for the very best journals'. [1133]

One respondent, in contrast, felt that the practice of providing evaluation criteria was 'a little looser than it used to be… Maybe ten years ago, a statement like 'Is the

paper clear and well written?' would have been broken up into a whole lot of subsections. [Now] the specificity… is perhaps missing. Sometimes people feel free to write their own interpretations without sticking to the criteria.' [1208]

There were also several other changes identified in the responses, including:

> There is an emerging hierarchy of journals, I think, encouraged by things like the ISI in the citation, people in the United States who classify journals as appropriate for their index or inappropriate, and a hierarchy of journals in terms of both the number of publications and the number of citations in particular journals. There has been a definite trend towards inclusion and exclusion of journals in terms of the worldwide hierarchy of journals. [1146]

Two respondents noted changes in the practice of editors. One felt that editors were 'much more interested in more pertinent advice. They like [reviewers] to be briefer and more to the point' [1185] and the other thought reviewers were 'urged more by editors to produce a 'quick response' to facilitate a quick reply to authors'. [3110]

Two education respondents noted changes in the research methods that journals now accept, noting that once 'unless you followed absolutely what were the dominant methodology and paradigm, there was no chance of getting anything else published', 'things have become a little bit more relaxed' now [1147] and 'most of the articles that I am being asked to review are more qualitative than quantitative; so there is more phenomenology, case study, ethnography, and so on.' [1151]

There are also changes in reviewers' behaviours. As a chemistry respondent noticed, there was

> … an increasing tendency for the reviewers to waive their anonymity in my field… people are little bit more relaxed about the process. There is a clarity aspect there that the reviewers are seen as contributing to the paper [and] often the authors acknowledge the reviewers by name occasionally for their contribution. [2189]

None of the respondents noted any change in the criteria for evaluation. Presumably this means no change has been witnessed. However, because 'criteria for evaluation' are a major focus of this study, the issue was probed further with the question: Have you noticed any change in the criteria for reviewing?

Five respondents answered 'yes' and provided brief explanation. Two felt the standard in publishing had dropped [1185] [3180]; one felt the standards were higher [1150]; and one with 25 years of reviewing experience felt the criteria she was given by journals were less specific [1202]. One noticed that 'there might be a bit more attention given to how well previous works are referenced', given that 'with the online databases it is easier for reviewers to check'. He also noted that some reviewers 'now show self-interest and ask for their own works to be cited.' [2189]

While most of the respondents sought some changes to peer review, most of them did not know how to achieve them. As one stated in relation to turn-around time, 'My experience is that it is often quite slow. I think there should be some incentive in order for this to be enforced, but I do not know how that could be done' [2173]. There were few specific strategies proposed by the respondents. The general desire for some change, yet apparent lack of wherewithal or determination to bring about change is another of the curious features associated with peer review which has been identified in the literature as well as here.

6.6 Learning to Review for Journals

Reviewers were invited to describe the way in which they learned to review. It was revealed that none of the respondents had received any formal training for reviewing. They were more likely to be 'thrown in and told to review [and] given no lesson [or] any instructions as how you go about it or whatever.' [1151] When reviewing for the first time, they would start by 'just doing it' [1151], using their 'common sense' [2166], and '[after] I did that one, then I did two, and then I did three. I guess I learn along the time.' [1163] The learning process followed a self-guided pattern, using sources such as related academic experience, reviews of their own publications, assistance from supervisors and colleagues, and journals' instructions.

'Adapting from related experience' appeared to be the major way of learning. Before their first invitation to review, they typically had accumulated experience in research and other academic tasks and had published [2215]. 'After you have been working in the area for a time, you have an idea of what's good science, and [reviewing] is really trying to reflect that' [3145]. They would apply the 'rigour [and] rules of engagement about how to produce a good journal article that you use to do your own… to [the reviewing of] other articles' [1221]. The process of accumulation could start as early as their doctoral study. As one recalled,

> I learned how to write for journals at the same time as I was learning how to review them, in that, my supervisor for my Masters and my PhD, and the first two papers I published were out of my Masters I was writing with her… I was learning how to write academically, along with people in this collaborative context, and I was learning how to evaluate the contributions we were each making to the paper as it was developing. [1255]

Some found the reviews they received on their own papers 'critical in forming [their] own approach to the process [1210],' from which they 'could see what reviewers were looking for' [1224] and 'learned by copying what other people did in the field' [3183]. Some referred to experiences of evaluating students' work, and viewed reviewing as 'the same sort of process as marking students' essays or examining a research thesis' [1154]. A respondent found commonalities in all kinds of assessment tasks, and reviewing was not 'much different frankly from a whole lot of other areas' where 'there are procedures and judgments and weighing up of different aspects of whatever is being put to you that is common across the range.' [1150]

The quote from respondent [1255] above alluded to the role that supervisors and colleagues played in the learning process, which served as the next major source. The first paper one reviewed was often passed on by their doctoral supervisors: 'my supervisor has been asked to review a journal article, and he asked me if I would do it for him… He gave me the article and said 'Go away and review that'.' [2177] This scenario was cited much more often by physics and chemistry respondents, for

example, 'in a tradition of academia, there is an expectation that you will train your post-doc people in the whole job; that includes reviewing papers.' [2129]

An education respondent recalled that 'when I got my first paper to review, I turned to my post-doc supervisor [who] has been an editor for very long time for several journals and I believe he gave me good advice.' [1163] Another one 'had a very good mentor in the school, and one of the first things he did for new academics was to get you into a situation where you would undertake peer review; this was very helpful and helped to set me up.' [1202]

Some experienced reviewers were aware of the needs of newer ones and consciously offered help. One described his intention to orientate junior academics into the task: 'I know I've had colleagues at university who've been asked to referee articles [who] have been hesitant about what to do', and 'I've given them, of course, anonymously copies of some typical reviews I have written about research articles, so as to give them confidence about refereeing papers themselves.' [1133]

In addition, three respondents acknowledged the helpfulness of journal guidelines, 'many journals will give their reviewers two or three pages of advice about what they are looking for, and I think you use that pretty carefully.' [1146]

Figure 6.2 summarises this self-guided learning process and the potential sources of external interventions. It shows that learning to review is primarily a self-guided process, with the engagement of limited mentoring assistance or explicit guidelines. For example, when students are asked by supervisors to undertake review, their instructions are 'not that detailed' [2141], so trial and error predominate. When receiving reviews on their own submissions, they have to decide which comments are worth following. When journal guidelines were provided, they still had to draw on their own interpretations of the requirements and 'try to understand everything and look for any problems and mistakes [in the paper] and to explain.' [3181]

At the end of this process, the once neophyte becomes 'an autonomous reviewer' with 'established standards' and 'an understanding of the process' [1147] who can then act as mentor for new reviewers. In this way, the process of preparing and

initiating the next generation of stewards (Walker, et al., 2008) continues, the criteria for evaluation are reinforced, and the skills, principles and perceptions of reviewing are inherited from one generation of reviewers to the other.

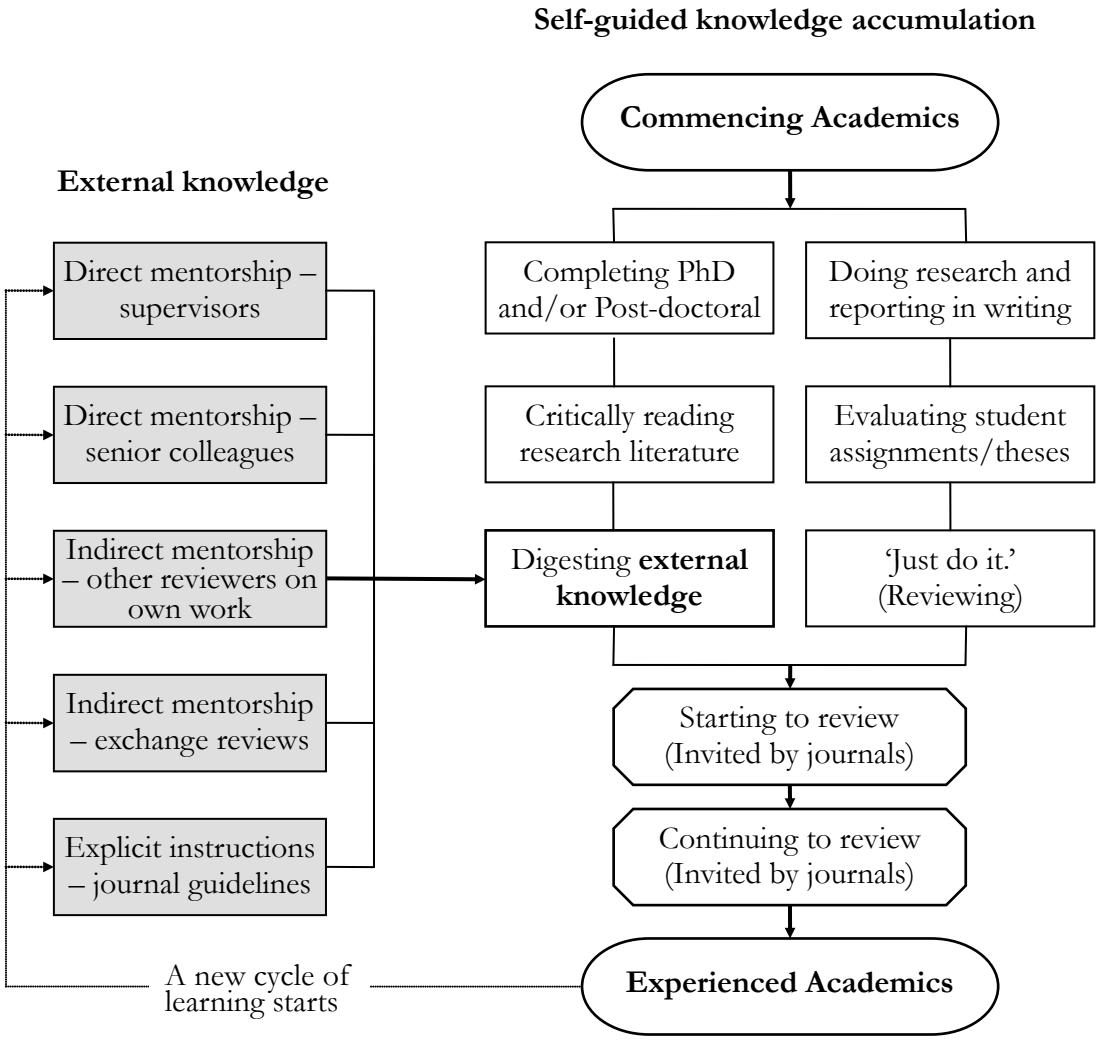


Figure 6.2 The Continuous Process of Learning to Review

This self-guided process of finding one's own path in academic growth is common in academic culture. However, will more organized, formal training be helpful in the generation of competent reviewers, especially given more good quality reviewers are now needed? After the first four interviews were conducted, a follow-up question was added to the protocol to explore the topic of formal training for reviewing.

6.7 Consideration of Formal Training for Reviewing

When the question 'Is formal training for new reviewers necessary?' was added, it was hypothesized that the respondents would support the idea of formal training, given that the first few interviewees all noted that when they were new reviewers, due to the absence of such training, they had to 'feel the way through' [1185]. It was also anticipated that some strategies of reviewer training would be offered.

However, of those who answered this question, 30 (75% of 40) said 'no' to formal training. Most of them asserted that it was not 'really necessary' [2107]. They felt that 'people can be good reviewers without formal training, just from experience' [2107]; 'if you have your own work reviewed, as years go on, you learn the methods of how to review' [2184]; 'reviewers are sensible, intelligent people and they can interpret that piece of paper without further training. It's just a waste of time.' [3103]; 'it's a part of the job and everyone knows.' [2138]

Some respondents perceived formalized training to be 'very unpractical' as 'it is difficult for journals to set up training programs… In an ideal world, a training package might be useful, [but] in a real world, pragmatically, it is too difficult' [1194] and neither 'any instruction [nor] online course would be particularly helpful' [1151]. Another questioned 'who could do that training' and whether there could be 'a generic form of training that would fit everything.' [1255]

Ten respondents said 'yes' to formal training. Some appeared to be quite passionate about it, by stating, 'Absolutely! All those things that I talked about don't come into your head in the every first instance.' [1202] and 'having some sort of mandatory process… would certainly strengthen the whole process' [1147]. One claimed that 'If journals want to maintain their standards, they could invest a little bit in the supporting and development process for reviewers' [1169] However, most of these indicated that they were 'not sure how you would go about doing that.' [1147]

Instead of endorsing formal training, more respondents offered alternative routes to training. Some suggested that it 'relies on the doctoral training exercise' [1126], 'all the formal training that is required comes along with the completion of your PhD,

doing a post-doc, and doing research in that field' [1163], and 'some sort of sharing of experiences and maybe workshop professional conferences that actually address the issue' [1208] should be incorporated into such periods. Some suggested training in the form of an on-the-job 'apprenticeship', 'where you had a new reviewer with one or two quite established reviewers' [1251].

A number of respondents explicitly trusted editors to identify the 'people who are well recognized in their research areas and have good track record as well as publications and research expertise' [1126]. Some noted that it was the editors, not formal training, that played the crucial role in producing good reviewers, 'It works best if the editors are very good, if they know the particular interests and specialties of the people who are on their review boards, and send them to the right people.' [1255]; and if they 'provide very clear, detailed, professional, unambiguous advice to reviewers about what they expect from the review process for their journals' [1154].

6.8 Confidence in Reviewing

'At what stage did you gain confidence in reviewing?' was also added to the initial interview protocol after the few interviewers: While a few respondents offered a time point when they felt confident, more offered indicators of 'being confident'. They perceived 'confidence' in three major ways: confidence of expertise in the specialized field of the work, confidence in their skill of critical analysis of the work, and confidence in their skill in composing a good review. Table 6.5 presents some quotations to illustrate the three types of confidence.

Some respondents indicated they had never felt really confident, while some others noted they had never felt short of confidence in the context of peer review. Quotes of these two kinds are provided in Table 6.6. As it reveals, the degree to which one feels confident is directly linked to a belief in how established they are in a field. This is not restricted to familiarity with a particular field, but involves experience in academic writing and evaluating academic writing. Such experience enables the reviewers to 'read critically' and judge whether a paper 'has been done in a scientific manner', even if the paper is 'in a particular region' that they are not very familiar with [3145].

Table 6.5 Participants' Perceptions of Confidence

Confidence	Examples
Expertise in a given field	The basic ingredient is that I am confident in my ability as a physicist and I have always been that. [3194]
	I think from being active researcher in the field, I have developed over the years a knowledge of the field and what is good science. [2189]
	I was confident enough from the beginning. I knew the sort of research was being published and I had ideas about what should be published and that's where I just went from. [1151]
Skills in critique	I felt my insights were getting at critical issues within articles. But I still come across articles I have to review where I feel very uncertain about, for example, where to start to critique the article. [1126]
	I think you also learn not to be too critical, to be rigorous but not too critical, and I think to recognize from a paper worthy things probably even though you do not agree with it, and you may feel there are some weaknesses in the argument. [1169]
	You begin to see a pattern and… you can quickly sort out which bits are good and which bits are poor. [1146]
Skills in composing a review	…how do you help the author say what it is you think they're trying to say, but not in a way which rewrites their article to your formula. It's got to have their voice, not yours, and in that respect [1146]
	I no longer typically try to get all the English right, and that sort of thing. I try to get to the core issues, especially for papers that I was recommending that the paper was got to be rejected. I do not go to lots of details. I just go 'This is not good enough because of X, Y and Z' and 'I suggest it be rejected'. [2153]

Although respondents were asked at what stage they felt confident to review, the 'stage' was not readily articulated. Several respondents did recall the rough point in time when they felt they became confident. One said, 'probably it took four or five years before I felt my insights were getting at critical issues within articles' [1126]; another noted, 'I think probably after I had done about a dozen reviews' [1146]; and a third: 'I suppose it probably took four years' [3183]. Overall, it would appear that academics need several years before they hit their stride as capable reviewers. This time might be shortened if there were training in place.

Table 6.6 Indicators of Confidence in Acting as a Reviewer

Do not question their confidence	Question their confidence
I got the confidence because I was asked to do it. I was recognized as a person whose opinion was valued and I took the task seriously. [1159]	Never [felt confident]. [Because] I hate the process of criticizing other people even though it has to be done. [1244]
I do not think I lack confidence, because I am a university lecturer and I mark students' work. I don't think that is a huge step from critiquing most students' work from critiquing other people's written work [1224]	I do not know if I have really. As for the question of correctness, I think that if you find a mistake, then this is probably discipline dependent… I do think I have become more or less confident on that. As for the question of judging the interests of research, it is more difficult to gain confidence of judging that. [3110]
I was confident enough from the beginning. I knew the sort of research was being published and I had ideas about what should be published and that's where I just went from. [1151]	No! Not at all! Especially when you get a paper that is not only not in your area but the area of science these days is very, very specialized… [If] you get articles that are just a little bit outside your particular area and many times, I am not at all confident that I have done the right thing. [3145]
I think my confidence has never changed. The basic ingredient is that I am confident in my ability as a physicist and I have always been that. [3103]	

Indications of confidence were evident in the responses. One marked it by the time when he rejected a paper for the first time, 'once I read the rejection letter from the editor… I knew I had done the right thing… That gave me confidence. I knew I had a role there. I felt comfortable at that point.' [1163] Another two indicated that they became less harsh as they were more able to identify merits in a paper, 'At the very early stage, you more tended to turn down a paper, but over the years you gradually learned what the appropriate standards were… you recognize any of the novel aspects' [2166] and 'you learn…to be rigorous but not too critical, and to recognize from a paper worthy things probably, even though you do not agree with it.' [1169]

A few respondents referred to the time they spent on a single review as an indicator. Generally, less time was needed as confidence grew, 'When I first reviewed, it took probably me the best part of the week to write the review. [Now] it probably takes me an hour… you learn some skills to review quickly and with confidence in the way which you comment on things' [1184]; '[for] some articles you pick up and you read

the first paragraph, and you know it doesn't go anywhere, and there are other articles you pick up and read the first paragraph and you're hooked… Those two things you gain confidence about pretty quickly.' [1146]

Some found the goal or approach they brought to reviewing had changed as they became more experienced. As one described, 'I do not spend as long a time to review a paper now as I did, and I no longer typically try to get all the English right. I try to get to the core issues, especially for papers that I was recommending that the paper was got to be rejected' because 'if the work is really inadequate in those areas, then generally speaking, it is really not my problem to fix it up; the authors need to be making more effort.' [2153]

'Decisiveness' for many was a sign of becoming confident in reviewing. Following a more or less self-guided initiation, they established personal reviewing patterns. By trial and error, and continuing confirmation, these patterns were consolidated and the reviewers eventually felt 'confident'.

6.9 Editors' Feedback to Reviewers

The last issue explored in this part of the interview is about how closely reviewers felt editors follow their advice. It reflects the extent to which editors rely on reviewers in the decision-making, as well as the level of feedback they provide to reviewers. The responses identified five kinds of editorial feedback, including an acknowledgement of receiving the review, a thank-you message, a notification of the editorial decision, editor's comments on the quality of the review, and an exchange of reviews between reviewers. The respondents demonstrated very different attitudes towards the necessity of providing feedback of different kinds.

The respondents viewed the first two kinds of feedback, acknowledgement of receipt and thank-you message, as a routine practice of journals. Most of them expected journals to send this message upon each review they completed, 'because it makes you feel better. It makes you more likely to do it next time!' [2107] Often this was the only feedback they received from editors, who 'didn't go into any detail about the quality of the review or anything like that.' [2107]

Although not all editors send out the 'thank-you' message, it did not appear to concern the reviewers very much. What really annoyed them was a lack of feedback on the reviewing outcomes. Nearly half of the reviewers indicated that, 'the main feedback I would like to receive is whether they publish the paper or not' [1147].

Whether and how editors notify reviewers of the final decisions varied widely across journals. Some editors informed reviewers about all papers they reviewed; some provided 'a list of the papers that you have reviewed, what your response was, and whether the paper was accepted or rejected' [2215]; and a journal had 'a particular web page that [reviewers] can go to and it gives you a history of your refereeing, all your comments that you have made and whether you rejected it, what your comments were, or whether you accepted that paper.' [2138]

However, most of the responding reviewers said they did not receive such feedback directly from the editor, 'I usually get… an email saying, 'Thank you for your review. It will be passed on.' But we don't usually get much more feedback than that. That's actually quite common.' [1146] Some read the journal and found the article they reviewed; some learned so when they 'recommended that the paper be revised and resubmitted' and it 'came back to me to review again' [1249]; and some just did not bother with the final outcomes because they were 'too busy to worry' [1159]. Some felt offended by the 'rudeness' of editors who did not inform them of the outcomes, who published the paper 'without any of your suggestions being commented upon' [3183].

Although the reviewers were not always informed of the final results, they generally felt positive about whether their advice was followed, 'by and large I feel satisfied that the editors at least acknowledged concerns that I've had and they always emphasized those concerns as priorities' [1126]. One noted that 'particularly for the reasonable to high quality research journals, the editors take the referees' comments pretty seriously' because 'it doesn't take long for editors to realize who their quality reviewers are and who aren't. I'm sure editors carefully select who they are going to send papers to because they know they are going to get reasonable comments.' [1133]

They generally expressed understanding for cases when their recommendation was not followed. 'Occasionally, when I turned down a paper, I saw it in the journal. This could mean that compared with the other referees, I was in the minority. It may also mean that the editors disagree with all the referees' [2166]. One respondent recognised that the rejection rate of a journal also played a role here, 'in terms of overall acceptance and rejection, particularly for those journals that are difficult to get into, often the editors don't follow my advice. They probably follow it in broad terms, except for rejecting more of them.' [2129]

The fourth kind of feedback, editors' comments on the reviews, was rarely provided, except for the occasional compliments applauding 'the detail and the constructive nature of the review comments' [1210]. Most of the respondents indicated this type of feedback was not necessary, and contended that the process was 'a one-way street' [1163] where they just did their 'little bit of the process' [1255], and 'on the whole it is better that the process is largely like that.' [2129]

Some perceived such feedback as a 'burden' for both editors and reviewers and it 'would be taking the process too far' [1151]. Many cited 'time pressure' as the reason: 'I cannot keep up with it if everyday there are comments' [2166]; 'I'm probably too busy to worry… If someone disagrees with my reviews, doesn't worry me in the slightest.' [1159] And as an editor-respondent asserted, 'We have so many more things to do. So I wouldn't give assessors and reviewers very much information or feedback unless I thought it was absolutely necessary.' [1185]

For those who provided poor reviews, some reviewers suggested they be excluded from the pool of reviewers 'If I felt somehow that they were being very personal about somebody, or they completely misunderstood the line of the article' [1185], yet criticism on all reviews was not necessary.

Several reviewers pointed out that some comments on review quality, and probably an anonymous exchange of reviews between reviewers, would be helpful particularly for new reviewers. Except for the purpose of training, the reviewers in general did not want any evaluation on their reviews. This question is the last one for this part of the interview. The next section analyzes data collected by the follow-up survey.

6.10 Themes Explored in the Follow-up Survey

The follow-up survey items were determined after the interviews were transcribed. The transcripts were read and notes were taken for intriguing convergent and divergent points. The notes were then synthesized; ideas and comments that were considered worth further investigation were incorporated into the follow-up survey.

When asked initially whether they had noticed any change or trend in peer review most of the respondents had responded 'no'. Similarly few commented on changes to the process. The researcher found this hard to believe given peer review attracts numerous criticisms for its flaws and wondered if it was somehow 'immune' to change because anything other than surface change may be too hard or too controversial. One of the more controversial areas, in education around that time, for example, involved setting criteria to judge the quality of research (Macnab & Thomas, 2007). Also there was an emerging concern expressed in the interviews about 'recognition' suggesting a departure from previous expectations, and a strand of comment recognizing the criticisms leveled at blind peer review. It was decided to pursue these themes further through email contact.

Three open-ended questions were constructed as follows:

1. Do you believe that peer review as a means to assess research quality can be improved further? What specifically can be improved?
2. In your experience, is the role of the reviewer held in high esteem by colleagues and your institution? Is the current level of recognition appropriate?
3. How highly do you value anonymous review? Would there be specific benefits if this practice were removed (as is beginning to happen in some e-journals)?

An email survey was distributed to the 44 interviewees; responses were received from 30 participants, or a response rate of 70 percent. The number of respondents was split evenly between education and physics and chemistry, as shown in Table 6.7.

The 'yes' and 'no' responses to the questions were presented in Appendix 14, divided by gender and academic title of the respondents.

Table 6.7 Demographics of the Follow-up Survey Participants

Discipline	Professor		A/Professor		S/Lecturer		Total	
	male	female	male	female	male	female	male	female
Education	2	1	4	1	2	5	8	7
Physics & Chemistry	6	1	0	0	8	0	14	1

6.10.1 Can peer review be improved?

The high response rate suggests the reviewers were interested in these questions, possibly because they could see how they related to their transcripts. In this round most of the respondents believed there was room for peer review to be improved. Only three responded 'no': one did not comment [2152], one simply wrote 'I am quite happy with current peer review process' [1133], and the third wrote 'peer review is at a steady state and effective. Technology helps in the process, but the basic structure is, in my opinion, sound.' [3145] It is interesting to note that there was some common features in the background of the three respondents: they were all male, senior in age and academic title, and all had served as editor and had acted as a reviewer for more than 10 years.

As for 'What specifically can be improved?', the major concerns were revealed to be the lack of qualified reviewers, the adequate interpretation of evaluation criteria, and feedback to reviewers; reviewer bias and misconduct; author identity being revealed to reviewers; and time wasted in resubmission to other journals after initial rejections. Due to the nature of email responses, short in length and improvised, it is reasonable to assume that what they wrote about were their strongest concerns. Table 6.8 presents a summary of the problems and the suggested solutions, in descending order of times that they were mentioned in the responses.

With regard to the lack of qualified reviewers, it was suggested that editors provide more guidance and on-going training for reviewers. The word 'clearer' was recurring

in the responses, which implied that there had been guidance but more detailed directions were needed.

Table 6.8 Participants' Suggested Solutions for Deficiencies in Peer Review

Problem	Solution
Lack of qualified reviewers	Guidance: clearer framework for reviewing activities
	Clearer criteria for reviewer selection
	On-the-job training
	Careful match of reviewers with manuscripts
	Recruit a mixture of new & experienced reviewers
	Better use of reviewer database
No common interpretation of evaluation criteria	Clearer instructions and criteria for reviewing for reviewers
	Use of standard reviewing form
	Clearer guidelines for authors
Lack of feedback to reviewers	Authors being able to respond to reviewers directly or through the editor
	Stronger editor leadership
	Exchanging reviews among reviewers
Reviewer bias & misconduct	Providing code of practice for reviewers
Revealing author identity	Nil
Time wasted in resubmission	A central reviewing panel for the discipline

The second common concern was the absence of well-understood, specified criteria for reviewing. As one wrote, 'I still have serious concerns about consistency in what reviewers are looking for. It seems that we don't talk enough about what 'quality' looks like.' [1224] Another wrote, 'There can be a wide range of standards applied by different referees. It would be helpful if the journal editors could establish and enforce some more specific guidelines, but these may be very hard to define' [3165]. The word 'clearer' was again recurring in the responses, indicating that 'usually criteria are given, but there is little attempt to develop common interpretations of their meaning, or to define the standards which are applied to them.' [1150] There was no evident recognition of controversy around criteria among this group.

The third issue was the lack of communication among editors, authors and reviewers. Several indicated that effective communication would be beneficial especially for authors, as an opportunity for them to respond to editorial decisions would help for example when 'reviewers are actually wrong in their comments, but the author being reviewed has no right of reply, the system breaks down completely' [1127]. Some called for stronger leadership and direct mentoring by the editor in this regard, that they 'should take up a more significant role than just passing on the comments of the reviewers to the authors' [2107].

In addition, one education respondent indicated his concern about issues caused by reviewers viewing authors as competitors, 'There are still too many reviewers who don't recognize vested interest or view submissions as competition rather than taking an objective position' [1163].

6.10.2 Is the role of the reviewer held in high esteem?

In response to the question: 'In your experience, is the role of the reviewer held in high esteem by academic colleagues and by your institution?', 12 said 'yes', two said 'yes' to recognition by colleagues, but 'no' to that by institutions, nine said 'no', and seven did not comment (see Appendix 14 for a breakdown of the responses).

In general, reviewing was perceived as being held in higher esteem by their colleagues than by the institution. Only three noted that their institution recognized reviewing in workload calculation and one noted that it was counted when assessing one's track records. Some attributed the lack of recognition to the practice of anonymous review, where reviewer identity was unknown to people other than the editor.

When asked: 'Is the current level of recognition appropriate?', 14 responded 'no', nine said 'yes', five said 'no need', and two did not comment. Nearly half of the reviewers expected a higher level of recognition, especially by counting reviewing as part of their academic workload. Some offered reasons for this. One asserted that, 'If the system is to work, it needs to be supported more openly and concretely' [1150]. Another took into consideration the broad context of publication and the knowledge economy,

> In our era where everything is counted for the sake of cost efficiency and productivity, it is plausible to suggest that the task of peer review be counted as part of the reviewer's workload, in the same way [as] publishing a refereed article or presenting a refereed conference paper or winning research funding. [1127]

Some reviewers thought reviewing did not require recognition, because 'The role of reviewer is an expectation of the duties of an academic. Everyone has to do it. You can't publish, obtain grants or supervise PhD students if you aren't prepared to review the work of others' [2215] and, if reviewers are paid, 'it changes a professional relationship and makes it an economic one' [1208].

6.10.3 The value of anonymous review

With regard to anonymous reviewing, 18 of the 30 reviewers noted that they valued it highly or very highly, five said not highly, four provided a 'yes and no' answer, one was 'not sure', and two did not answer (Appendix 14). The second part of this question asked: 'Whether there are specific benefits if this practice were removed?' Twenty responded 'no', six thought that it was up to the reviewers to choose to be anonymous or not, two did not comment, and only two said 'yes'. Several reviewers emphasised the benefit of anonymous review: it not only promotes objective and honest reviewing without fear of, for example, 'running into the author at a conference' [1224], but also prevents conflict of interest between reviewers and authors, 'especially where one of the parties has greater influence, academically or institutionally, than the other and could use his or her position to damage the other's career in some way, if only by barring publication or funding.' [1127] It can also protect young reviewers when they criticize the work of established academics [2215].

The reviewers who offered a 'yes-and-no' response agreed with others on the benefits of anonymous review, but believed that its advantages would be outweighed by its disadvantages. The six who wrote 'it depends' noted that being anonymous or not should always be at the discretion of the reviewer. The two 'yes' responses were both from education; one wrote, 'If you believe what you write about another's work, you should be happy to have your name on it' [1202]; the other did not offer a reason.

6.11 Conclusions and Discussion

This section brings together findings of the interviews and surveys. Issues that will be discussed include: academics' faith in peer review and tolerance of its deficiencies, recognition of reviewers' contribution, anonymous review, and reviewer training.

In general, the respondents experienced high expectations of, and satisfaction with, peer review, as well as a strong faith in it. It may well be that those who volunteered were of that persuasion, but it has already been demonstrated elsewhere that the contradictory position of recognising flaws but accepting the system of peer review is a general tendency among academics.

'Faith' refers to a deep conviction of fidelity and integrity, often regardless of lack of, or even in the face of contrary, evidence. The sense of deep conviction recurred in the responses. The respondents maintained that peer review was essential to the academy, especially for its contribution of ensuring high quality in academic research and publication. Although most of them were really critical towards peer review, they often reiterated its essentiality right after they made a critical remark or noted that the problems were 'related to individual cases which relates to the behavior of individual referees' [2153] while the system in general worked well. Similarly, while many admitted that there was room for change, they indicated that to make such changes were not easy, and no one had suggested abandoning peer review.

Many respondents took the last interview and/or survey question for free comments as an opportunity to express their faith in peer review, making statements such as:

> It is crucial for developing a knowledge base that's sound and robust, for the time being, something to base practice on. [1221]
>
> Peer review is essential to the whole process of academic research, so one has to support it. [1126]
>
> Peer review is the best system, so whatever the shortcomings we should try to improve it. But you can't dispense with it; there is no other way; we must have a peer review process. [3180]

> The editors and reviewers are generous people who are selected for their judgment, who can judge conflicts of interest and make sound decisions. [2215]

Aligned with their faith in peer review, the respondents demonstrated a very high level of tolerance of its deficiencies. Reviewers' tolerance may be bolstered by the voluntary nature of the process and a degree of temerity that alternatives may be worse, especially in a system under pressure: 'Most academics act with a lot of integrity and commitment to the review process… I cannot see anything else can do the same job' [1169]; and peer review 'is driven by the scientific community volunteering their time, without getting paid to do reviewing of manuscripts', so it shall not be replaced unless 'someone can come up with a system that works better as a proper control of practising scientists.' [2139]

Several respondents expressed their tolerance by comparing peer review to 'democracy', 'which has been defined as the worst political system of government known, except for all of the others' [1244], in the sense that 'it's not perfect, but it's the best we have.' [3181] In addition, senior academics appeared to be more tolerant than more junior ones. As some stated, this would be because established academics were less vulnerable to the negative effects caused by the deficiencies.

If one analogises scholarly submissions to traffic on the way to publication, peer review is the solo 'traffic policeman', and reviewers are the traffic lights and cameras who help to control and stratify the traffic and punish the poorly behaved. Today, the 'traffic' has become increasingly heavy and the 'drivers' (authors) are all very keen and impatient to reach the destination (publication). As such, it becomes almost unacceptable, if not impossible, to pull the 'traffic policeman' off the frontline for a thorough 'health check' or at least treatment for existing wounds, not to mention replacement. For the same reason some 'glitches' and breakage in the infrastructure are noted, but left unattended.

This study revealed that the responding reviewers were generally satisfied with the esteem they receive from their colleagues generated by peer review and the literature does outline the tangible gains from involvement through networking, and gaining

experience (for example, Braun & Bujdosó, 1983). The initial survey also revealed that reviewers treated peer review as a part of professional obligation and in generosity of spirit: 'they will rarely refuse the invitation and they have always done a professional job… I can't ever remember being disappointed with the effort the reviewers appeared to have put into the task.' [3102].

However, there was a strong call for the contribution of reviewers to be recognized appropriately in a more explicit manner. In a world of intensifying quality assurance, and accountability where academic workloads and activities are now quantified, this is not entirely surprising, but it does call into question how long a voluntary system can endure.

The issue of overloading qualified reviewers, and the pressure on all academics, has exacerbated the need for the appropriate recognition of reviewers. After all, reviewers are also required to write papers in the time available to them, so unless peer review is 'recognised, people will just start saying no' and 'payment doesn't solve the problem of time, and that's the big problem for most of us' [1210]. As many of the participating reviewers believed, a legitimate recognition by employing institutions will prevent the 'pool of good reviewers' from shrinking (too much) if they find their goodwill is reciprocated.

Another area of concern is related to the criteria of evaluation. A number of the respondents cited 'establishing a standard' as a major contribution of peer review, as if the standard is defined by peer review. However, at present, peer review identifies high quality research on the basis of quality defined by itself, yet 'what good quality is has rarely been discussed' [1208]. It is pertinent to note here that the whole area of research assessment and how judgments are made, from PhD examination (Kiley & Wisker, 2009; Lovat et al., 2009; Lovitts, 2007; Trafford, 2003) to grant peer review (Guetzkow et al., 2004; Jayasinghe et al., 2001; Lamont, 2009) is still very much a new area of endeavour that grew to prominence only in the last decade. Highlighted in reviewers' responses in this study is a call for developing clearer criteria that specify what 'quality' is for journal publication. As the respondents are among the most experienced, this is a noteworthy finding.

A better understanding of the explicit criteria applied by those with experience is clearly a useful way forward. However, there are a number of issues affecting the development of such explication. First, what qualities need to be defined? The literature offers a wide range of features that characterize an 'ideal' paper, and all those who involve in research and scholarly writing seem to have their understanding of standards (Frank, 1996; Sigelman & Whicker, 1987). However, what reviewers typically receive are papers flawed in one way or another. Their task is not only to detect flaws but to judge, despite the flaws, which ones are worth publishing and which ones are not, or, where the lines are drawn to differentiate quality.

The second issue is about how explicit the criteria should be. The literature shows that some criteria are understood better than others, such as methodological errors and language usage. In contrast, 'originality' is an example of criteria heavily weighted in editorial judgment, but poorly defined. An interview-based study with 67 grant assessment panelists revealed that between and within two related fields, humanities and the social sciences, there was considerable variation in the understanding of originality (Guetzkow et al., 2004) and what cutting edge is (Forbes, 2003). So there is also a need to obtain a fine-grained explication of all the key criteria.

The third issue is about how generic such criteria should be. As the analysis of journal policies (Chapter 4) shows, many journals have established policies for the scope, style and standard of publication, and journals produced especially by major learned societies or commercial publishers usually share well-established policies. What is lacking, according to the respondents, is a level of 'uniformity' within a field where values are shared by all editors, reviewers, and authors in that field.

The above three concerns were further explored in the next part of the interview, as well as by the analysis of review reports.

In addition, the study revealed that most of the participating reviewers were committed to an anonymous system; only a few suggested that releasing reviewer

identity to authors would make the process fairer. Academic attitudes toward the value of anonymity are influenced by two clear concerns. First there is a concern that if the practice is abandoned, the disadvantages will outweigh the advantages, because there are more known problems in a non-anonymous system.

Second, it is a matter of who will benefit from an open system. As the literature (for example, Kearney & Freda, 2005) and this study has revealed, academics feel authors deserve the right to know who reviewed their work and most believe that anonymous review protects reviewers from potential abuse and revenge. Even for those who would like to see an open system to be applied to some extent, they maintain that the right to choose to be anonymous or not remained with the reviewers. In particular, they are not very positive about 'open peer review' mostly trialled by electronic journals, as they tend to deem these as of lower quality than traditional print journals.

The last issue being covered by the first part of the interview is concerning the training of (new) reviewers. The literature shows that the effect of certain short-term or distance package-based training on improving review quality was limited (Callaham & Tercier, 2007b; Schroter et al., 2004). The respondents in this study did not support the idea of formal training, perceiving such training to be either unnecessary or impractical to implement. Instead, a few of them proposed several forms of long-term, on-the-job training, such as: incorporating knowledge of peer review in the training of research students, encouraging communications between reviewers and authors and among reviewers and exchanging reviews between reviewers (via the editor and anonymously), and recruiting reviewing panels of both experienced and newer reviewers.

In the next chapter, the interview responses regarding criteria for the evaluation of manuscripts will be discussed, and these responses will then be compared with what the reviewers actually wrote in their reviews (Chapter 8).

CHAPTER 7: WHERE THE LINE IS DRAWN IN REVIEWING DECISIONS

7.0 Introduction

Kassirer and Campion (1994, p. 98) claim that, in the evaluation of a paper, there is a 'rejection threshold' at which the 'cumulative weight' of the paper's flaws 'tips the scales toward rejection'. The authors focused on flaws in articles in the biomedical sciences. The four major categories of flaws in order from most to least important were: design, presentation, interpretation, and overall importance of the research. McKercher et al. (2007, p. 468) identified that, in hospitality and tourism, the most common reasons for rejections were flaws in method, failure to elucidate significance, poor writing style, and weak literature.

The literature offered a number of reasons for rejection and acceptance of a paper (for example, Bakanic et al., 1989; Henson, 1993; Noble, 1989), and the analysis of editorials (Chapter 4) revealed the criteria emphasised by editors. However, rarely has the literature provided further detail on 'thresholds' or explored decisions relating to papers that contain flaws of different types.

In response to this gap, the second part of each interview focused specifically on what reviewers perceived would lead to different reviewing decisions and recommendations. The interview questions covered the following:

- What features would always lead you to recommend rejection of a paper?
- What are the essential features necessary for you to accept a paper?
- What differentiates a high quality paper from a marginally acceptable one?
- Thinking about the best research paper you have ever reviewed, what characteristics of the research impressed you the most?

These questions were constructed in a way to tease out finer distinctions (or 'lines') between qualities that led to different reviewing decisions across the spectrum from

rejection and marginally or potentially acceptable papers through to high quality and the 'best'. The sections in this chapter are structured in accord with this sequence. Also, as some respondents also served as ARC assessors, they were invited to compare their considerations in the evaluation of research projects with those for journal papers,

- What are the similarities between the assessment of a journal paper and the assessment of an ARC project, in particular with reference to the research?
- What are the differences between the assessment of a journal paper and the assessment of an ARC project, in particular with reference to the research?

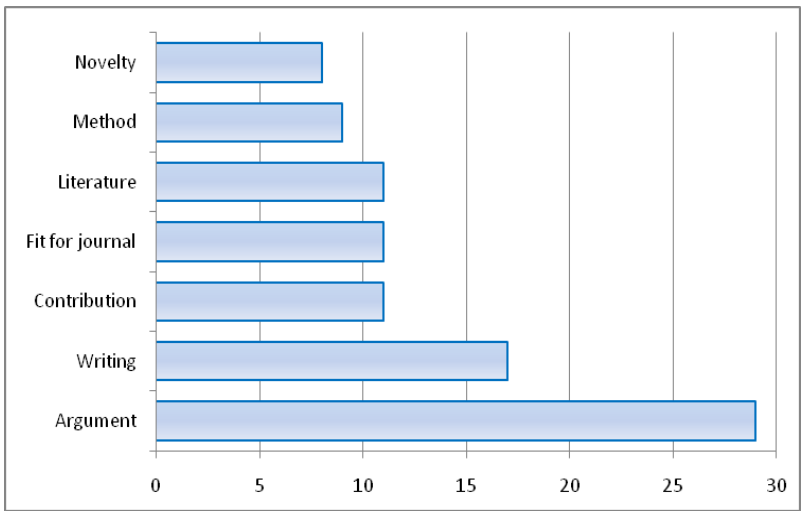
The responses to these two questions are discussed in the final section of the chapter.

7.1 Reasons for Recommending Rejection of a Manuscript

On a typical reviewing form, one recommendation that reviewers can nominate is 'Reject'. The participating reviewers were invited to nominate features that would always lead them to reject a paper. Some reviewers cited a single reason, while others offered more. The reasons fell into seven content categories with no outlier: writing, argument, literature review, research design and methods, contribution, fit for journal, and novelty.

Figure 7.1 presents the number of responses that fell into each category and these in turn give some sense of the weight of emphasis in the responses by the group. A category was counted only once for each respondent. As it shows, the presence of flaws in the *argument* was mostly likely to attract outright rejection, followed by flaws in *writing*, while problems related to *novelty* and *method* were the least cited. This is not consistent with the emphases identified by the studies using similar methods (Kassirer & Campion, 1994; McKercher et al., 2007). This is probably because the question asked in this study is specifically about '<u>always</u> leading to rejection'.

Figure 7.1 Reasons for Rejection: Count of Responses by Category



The reviewers were most likely to recommend an outright rejection if they found the argument was substandard. The evaluation of argument focuses on two aspects: clarity and logic in the argument, and whether 'the evidence supports the argument' [1169]. The reviewers would reject a paper outright if it is 'convoluted', 'it just reports without arguing, if there's no point being made, if the paper meanders between themes' [1202], if 'the argument is loose and not clearly and precisely made' [1150] which makes the paper 'hard to follow' [1249] or for readers to see 'the point of it.' [1169]

An argument in which 'the evidence provided doesn't warrant the conclusions drawn' [1147] can also lead to outright rejection. What reviewers looked for was a sense of 'trust' [1244] in the conclusion. They would reject a paper if:

> The evidence provided doesn't warrant the conclusions drawn. [1147]
>
> [It reports] results from a small study using half a dozen kids in a particular location but concludes as school children in Australia… generally. [1185]
>
> [Authors make] claims according to what they want to believe. [1202]
>
> [There is] a complete misinterpretation of data. [1169]

The second major reason for outright rejection was *poor writing*, including poor spelling, punctuation, sentence structure, grammar, and poor usage of scholarly style. Some reviewers questioned the author's attitude if there was poor writing,

> If someone sends in something that… has grammatical errors all the way through it, then you think, 'Oh gosh, they should have not sent in a first draft.' It does reflect how carefully they have approached the whole of the task. So that annoys me. [1255]
>
> If a person has been very sloppy in writing a paper, they have probably been very sloppy in doing the chemistry. [2138]

Some respondents expressed sympathy for authors from a non-English speaking background. As one noted, he would 'not outright reject the paper' written by such an author, but would suggest it 'be edited by an English speaking editor or the author be provided with some other English aid.' [1163] However, most of the reviewers emphasized that when a paper reaches the stage of submission, 'as we are all writing in English, it's got to be at least reasonably soundly put together… It doesn't have to be perfect, but it does have to be readable.' [1185].

Compared to argument and writing, three other reasons for outright rejection were noted by many but did not call forth much explanation. The reasons related to *contribution, literature review* and *fit for journal*, as these quotations illustrate:

> If it did not contribute to a better understanding of the field… [1126]
>
> [There was] a lack of substantial literature review that justifies and conceptualizes the research question. [1194]
>
> Misunderstanding of key literature… [1146]
>
> If the topic of the paper does not fit the usual focus of the journal… [1227]

While reviewers from the two fields did not differ in any obvious way in the elaboration of the aforementioned five reasons, they did so when citing 'novelty' and 'research design and methods'. About half of the physics and chemistry reviewers cited *lack of novelty* as <u>the</u> reason for outright rejection, while only one education reviewer cited it as <u>the</u> reason. This reviewer was from science education, a field

where the assessment of research could be influenced by the tradition of the science. The respondents indicated that they would reject a paper if:

> There's already been another 15 articles written on the topic before and what [the authors have] done doesn't add anything different to it. [1226]
>
> It is reasonable work but not at the cutting-edge. [2141]
>
> The science is… in an old fashioned, well-worn area. [2169]

Disciplinary variation was also found in the responses about *research design and methods*. Education reviewers tended to make broad statements, for example, the author 'clearly does not have a formulated methodology and does not follow it' [1151] and 'the research was not very rigorous, if that was low level research' [1224], while physics and chemistry reviewers offered more explicit descriptions, for example, 'the correctness of the results is the main criteria for outright rejection' [3110], 'the methodology is not correct' [2189], and 'it looks like they haven't done the science right, there are wrong statements, just obviously wrong, that will be rejected.' [3145]

While reviewers noted that papers with flaws such as 'wrong science' [3165] must be rejected, they perceived some other flaws as more or less remediable. Some noted that they 'would always try to encourage somebody to revise the article for publication if there is anything that is worthy to be revised' [1194]. It would appear that reviewers distinguished between flaws that a paper could have – some attract outright rejection while others lead to a recommendation of revision.

Do reviewers draw this distinction between the same sorts of flaws? And how consistently do reviewers go about this? Answers to these questions became clear as the analysis continued. The next they were asked was 'what were the essential features for you to accept a paper?'.

7.2 The Line of Remedy

A paper needs to satisfy a number of requirements to avoid outright rejection. Such a paper will typically still contain a few flaws, but the flaws are remediable, as

perceived by the reviewer. The respondents provided detail of such requirements, which fell into the same seven categories as for rejection. This is not surprising that the categories mirror one another, given the nature of the two questions is to seek the features that distinguish between outright rejection and potential acceptance.

Figure 7.2 Reasons for Marginal Acceptance: Count of Responses by Category

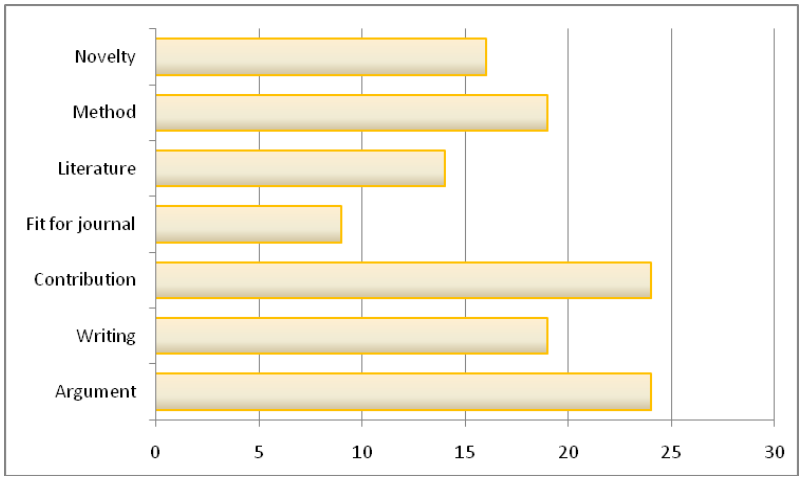


Figure 7.2 shows the number of responses in those categories. Here the pattern is somewhat different. To be a potentially acceptable paper, it has to present a sound argument and clear contribution, since 'there's no point putting something out there if people find it difficult or not helpful in terms of reading it' [1202]. However, how do all the quality features configure in the way reviewers perceive they differentiate between rejection and potential acceptance?

Figure 7.3 Comparisons between Reasons for Rejection and Marginal Acceptance

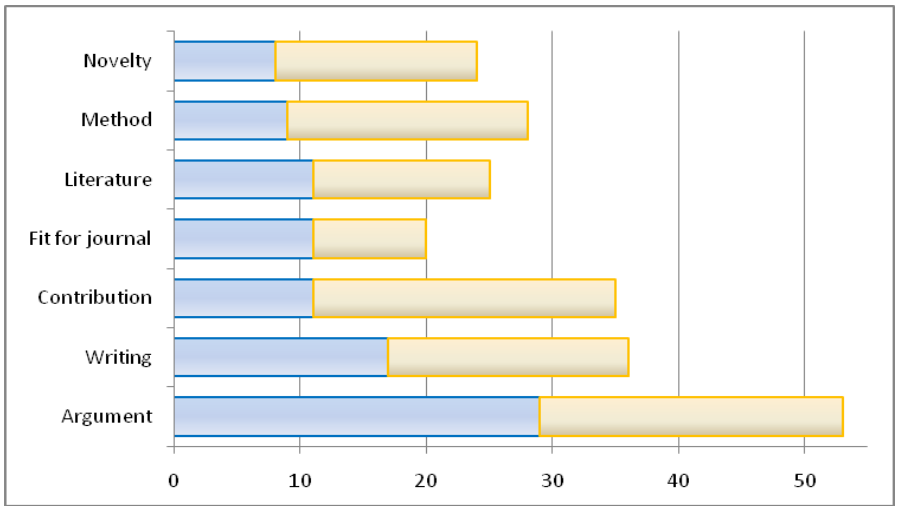


Figure 7.3 presents a tentative comparison between reviewers' considerations regarding rejection and recommending revisions, as revealed in their responses, by merging the bars in Figures 7.1 and 7.2; the merged bars suggest the relative 'power' of the features used to make the distinction, although it was acknowledged that the decisions were complex.

As Figure 7.3 shows, four of the features, *argument, writing, literature review*, and *fit for journal*, were cited at a similar level in the responses regarding rejection and potential acceptance. Table 7.1 presents some quotes to illustrate the threshold between being potentially acceptable and completely unacceptable for the first three features. These features shared at least two commonalities: the threshold was marked by the 'extent' to which a paper satisfied the quality perceived by reviewers, rather than a dichotomous judgment of yes or no; and there was always room for improvement. The researcher defined problems related to these features as 'remediable problems' (Group A problems).

There are also very pronounced differences dividing acceptable and unacceptable quality within the same categories of features. As Table 7.1 shows, manuscripts with acceptable quality are tight and credible, showing good communicative competence, and capable and informed treatment of literature. They are distinguished clearly from manuscripts of unacceptable quality where the argument is loose and overblown, the writing is incompetent and sloppy, and the review characterized by a lack of substance and poor decisions.

As for the fourth feature, *fit for journal*, a paper could be 'a great article but [sent] to the wrong journal' [1194] and rejected, but still be accepted by another journal. The researcher defined it as a relatively remediable problem.

For a paper with remediable problems, if reviewers found 'there is anything that is worthy to be revised' [1169], they would 'be collegial' [1194] and 'accept the paper [and] just recommend improvement' [1163].

If problems are remediable they may not contribute to outright rejection, however problems related to *research design and methods, contribution,* and *novelty* were viewed

as irremediable or 'not being able to be salvaged in its present form'. Flaws in these are deep and structural, that is, 'you can't just do some minor modifications; the whole thing is false and it has to be rejected' [3180]. 'Irremediable problems' are also defined in this thesis as Group B problems, that is, Group A are remediable and Group B are not. (This briefer wording becomes useful in later discussions.)

Table 7.1 Distinctions between Acceptable and Unacceptable Manuscripts

Feature	Example	
	Acceptable quality	**Unacceptable quality**
Argument	It has an interesting and carefully constructed argument [1154]	The argument is loose and not clearly and precisely made [1150]
	It does not try to achieve too much in one manuscript, it is clear what the goals are [1126]	If people make broad, sweeping generalizations or if they are drawing conclusions that are not founded on a sufficiently well established premise [1147]
	The paper has to contain enough evidence for the reader to have confidence in the conclusions that are drawn [1249]	Exaggerate the significance of the paper and say that its purpose is actually greater or different from what it turns out to be [1194]
Writing	They say what they want to say clearly and simply and as briefly as they can [1146]	It is badly presented, obscure, ambiguous, unclear, and hard to understand [1218]
	A well written paper, makes good sense, no flowery wording [2138]	Indeed very poor spelling, grammar, punctuation, sentence structure [1242]
Literature review	Strong critically informed grasp of existing research knowledge about the topic [1127]	It needs to contribute to my knowledge and every reader's knowledge, not just to the citation count of the writer. [1255]
	It ties into already available literature and research that has been completed [1147]	They hadn't been very clear in the literature review about what the focus of their epistemology was and what their research was about [1221]
	The referencing is always absolutely spot on; that one knows why they're in there [1202]	Lack of substantial literature review that justifies and conceptualizes the research question [1126]

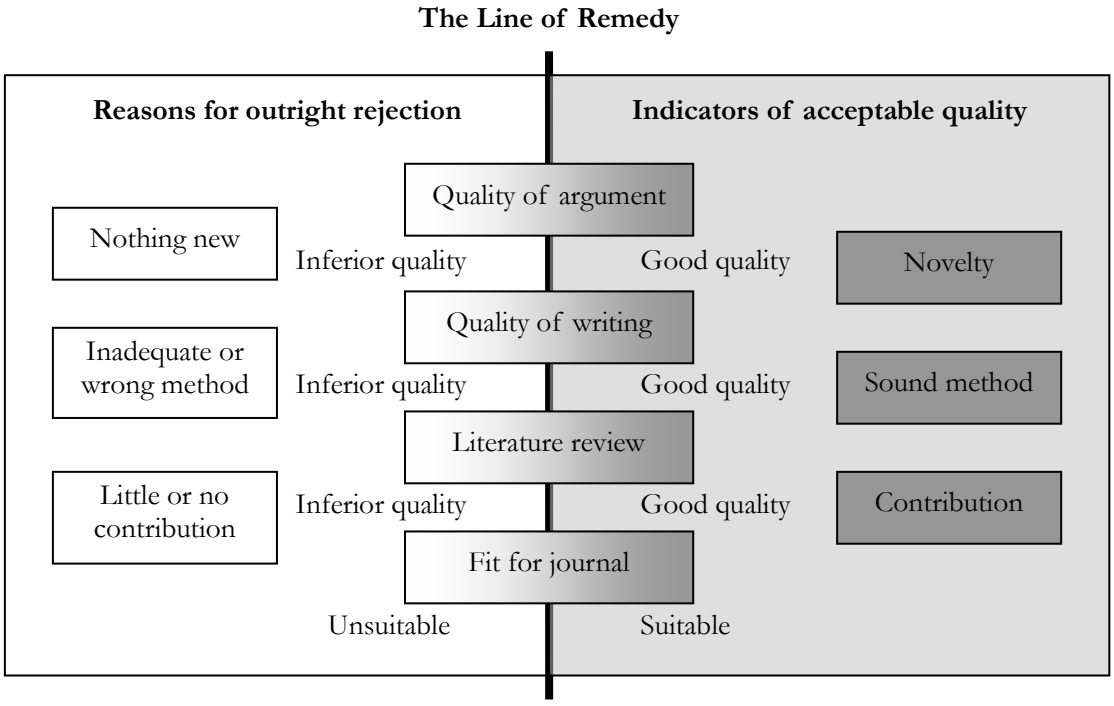
Papers with irremediable problems were considered as 'badly flawed' [3183]. For example, 'if the whole paper was based on a false assumption or a false interpretation' [3183] or if it 'has a fatal flaw in the methodology or the analysis is poorly undertaken or the quality of the data is poor or the conceptualization is

weak… sometimes even with extra work… there is nothing you can do. You just reject it.' [1159]

Specifically, with regard to 'research design and method', the reviewers expected that: 'it is technically of a good quality' [2107], 'the methodology must be clear and logical' [2189], and 'the experimental detail is appropriate for the particular topic' [2129].

For 'novelty', it was expected that the paper should, 'at least in a small way, offer a new way of reflecting on things' [3102] which 'might be just a different form of data that [they] have never seen before' [1255] 'that catch your interest' [3165]. While for 'contribution', it was expected that the work 'has to be a significant and useful advance on existing knowledge' [3103] and that 'substantially at least some new ground has been broken' [1151]. The reviewers highlighted the essential nature of the three features by using the word 'must' repeatedly. Drawing together the emphases, explanations, and qualifications in the responses provides a more nuanced picture of decision making. As Kassirer and Campion (1994, p. 98) note, there is a 'cumulative weight' of flaws, and it would appear that poor argument may well 'tip the scale' if other flaws are detected in a marginal paper.

Figure 7.4 The Line of Remedy



There is a strong indication that, when reviewers evaluate a paper, they draw a clear line between what must be there for acceptance and what leads to outright rejection. This is the *Line of Remedy*. Figure 7.4 illustrates it with a thick line in the middle of the diagram. The features in Group A sit across the line, showing that they are gauged by extent; Group B features are located on both sides of the line, indicating they are differentiated dichotomously. The *Line of Remedy* shows not only the feasibility of remediation, but the likelihood of acceptance, where papers falling into the right grid are promising while those on the left have little chance.

Apart from the seven features, the data also revealed an additional feature for acceptance, namely, *being interesting*. This feature did not clearly fall into any of the typical groups of criteria, as whether a paper appealed to reviewers as 'interesting' depended very much on their 'subjective' [2177] judgment.

A few of the respondents offered further explanation. A chemist defined 'being interesting' as 'a measure' of 'whether the work is likely to attract the attention of researchers in that particular area' and asserted that 'generally speaking, the bigger the circulation of the journal, the harder it is to get into it, and usually the higher the interest' [2177]. Another noted that a paper submitted to 'a very high ranking journal' had to

> … have some good impact or novelty associated with the chemistry. If it's fairly mundane chemistry, it's new chemistry but nothing essentially exciting about it, then you might reject it from the point of view that it doesn't have the novelty about it. [2184]

An interesting paper is 'basically what is wanted by publishers' [2153]. This is probably because, as the data suggest, the feature of being interesting is intertwined with the paper's capability of attracting readers and so boosting a journal's circulation. It uncovers a corner of the iceberg of a new level of quality – one that is higher than marginally acceptable and signifies the potential of the paper for long-term impact.

7.3 The Line of Excellence

Apart from drawing the Line of Remedy, reviewers also make distinctions between marginally acceptable quality and high quality, defined in this thesis as the Line of Excellence. The respondents provided insights into characteristics that set 'a paper apart' from other 'normal papers' [1126]. As one respondent noted, such characteristics were 'the most difficult area [to study] because it's a grey area' [1147]. The reviewers drew on two distinctive frameworks to distinguish high quality from acceptable quality.

The first framework nominates a mix of all the features cited previously but 'in better quality' [1194]. As one explained, 'I use a kind of matrix in my head, 'To what degree?' If it barely achieves all of those things, it's marginal. If it achieves those to a high quality all the way, then it's a very high quality paper.' [1202] Another emphasised coherence, in that 'when all of those characteristics that I have just described are there', they 'all come together to form a coherent argument that is well structured, well researched, and in a way, holds your attention' [1249]. This framework was mentioned much more often by education respondents.

The second framework identifies a single feature to distinguish high quality from average. Although other features also contribute to the 'cumulative weight' of a paper, this feature plays a prominent role. The reviewers consistently recognized 'novelty' as the single feature. Many also cited 'well-written argument' as a crucial feature in making the novelty of a work easy to understand. The following quotations illustrate the two features.

Novelty

> The extent of interestingness depends on the research problem: that people are creative about the problems they are exploring and they look for different theoretical frameworks that inform that problem or different methodology to collect information about it. [1126]
>
> I think it must have something to do with the significance. I mean I have seen something that is interesting, something that is really of high quality,

> something which is really breaking new ground in a field, makes you think in a way as a reviewer you have never thought of previously. [1169]
>
> If someone's… done something from a completely different point of view, that goes against normal chemical knowledge and they've done it and it's worked and you think, 'Wow, that's really exciting stuff!' [2138]
>
> A high quality paper leaves me thinking: Yes, that is right! I hadn't thought of these things before, or hadn't thought of these things in this way before, but I am convinced now that this is what I should think, and do. The highest quality work succeeds in changing me. [2177]

Argument

> The writers have been really careful about what type of procedures they used and they have tried to justify the use of those procedures in relation to their research questions and all of that hangs very nicely with how they've come up with their findings. Their evidence and their data seem to fit with their findings and they haven't gone overboard with their findings and they've been careful in relation to those. [1133]
>
> The really good papers have powerful, absorbing introductions and good conclusions. [1185]
>
> How open it is to interpretation, is the experiment justified, is there one interpretation or is there other interpretation? [2107]
>
> These high quality papers are easy to read, they provide all of the background you need to read the paper, they educate, they present. [3181]

While the respondents all prioritized novelty and well-written argument as the key to detect papers of high quality, compared to the framework of 'all and better', the 'well-argued novelty' framework was much more often cited by physics and chemistry respondents.

7.4 The Line of Impact

The sequence of the interview took the respondents more deeply into good quality, specifically to explore reviewers' conceptions of the next level of quality, namely, the 'best' papers. When this question was asked, only very few of the respondents instantly thought of a paper, yet most of them offered some description of what they perceived such a paper could be like.

Most respondents indicated that they had not read or reviewed a paper that they could call the best, and they often said so with a tone of regret, noting that reading 'something that you thought 'Wow!' hasn't happened in my career as a reviewer' [1133], 'none comes to my mind immediately. I have reviewed some quite good papers, but none sort of jetting to my mind as the best' [1224], 'a paper that I have finished reading and thought, 'Wow! That was really something!" [3102] or rather, 'Quite often I think I am getting papers [at] the other end for some reason.' [2152]

Their tone suggested that the reviewers were longing for something they could be 'wowed' by. Although not many of them had the 'fortunate' experience [1133] of reading a paper that was really surprising and admirable, they all seemed to have mapped out what such a paper should be like. The responses did not add any new feature of quality to those already discussed. Rather, they reinforced that a top quality paper should satisfy all the features 'in a very high degree' [1151]. Two quotes offer a nice summary of this 'totality' and 'superiority':

> The combination of topicality in terms of hitting an issue that's very topical at that point in time, and of relevance to the journal's audience, also being sort of a tick on all those quality measures… [It] addresses the conceptual arguments; it has to be methodologically sound, has engaged with relevant literature and been writing in an engaging and creative way. So it's a combination of factors. [1210]

> The features of the research included: simplicity: there was little use of obscure jargon: the approach was immediately obvious and compelling in a common sense, that's-exactly-the-right-way manner; practicality: the data gathered and analyzed was obviously the right data to gather and it was

> discussed in a direct and compelling way; connectedness: the research was clearly well informed and critically analyzed in terms of the existing background knowledge on that topic. Good research is convincing. [1127]

The responses also confirmed the superior role that 'novelty' and 'well-written argument' played in making a high quality paper. This paper did not only report a contribution that was new, no one else had been able to achieve before, and significant, but also had to underpin its contribution by strong, evidence-based, and well-written argument which was 'very clear, very erudite, and very simple, so you know exactly what the points are being made' [1202], and that 'the quotations in it were appropriate and you could see why they had been selected. It was an effortless read really.' [1249]

Some reviewers indicated that papers of a really high quality often came from authors who have a 'really good understanding of the phenomena that they were discussing' which was 'gained through years of their own work' [2173], and who have a 'good strong vocabulary for the field' [1249], as opposed to someone who just writes something 'off those well read books' [2173] or just 'for the sake of getting a journal article published to add to their research activity score' [1154].

It should be noted that, when talking about the best paper, the tone of the reviewers became much more subjective and vivid than it was for rejection and acceptance. More first person expressions were used to articulate personal perceptions and judgment of the 'best'. Quotations to illustrate this are presented below:

> For me personally, it's readability. If I'm reading a paper I want it to be an enjoyable experience. I don't want it to be like pulling teeth, where you have to go through all this really hard stuff to work out what the person is trying to say. An enjoyable paper is a good paper… It's the sort of paper that I really don't want to put down… [They] are just so intrinsically exciting that I want to find out more about what's going on in that particular area. [1147]

> In essence, it's intellectually challenging and exciting and I want to read it, I want to learn from it. [1154]

> They just made me think, 'That is not impossible.' The results they have shown me were unexpected… Before you read the paper, you assume you are familiar with this field. And the paper shows you something new. It takes your breath away and you think, 'Wow, they can do that!' [1163]
>
> For me, style is very important and I can think of a paper I read a couple of years ago, and my reviewer's comment to the editor was, 'Publish this paper as soon as possible.' It was like a good novel: I really enjoyed reading it. I couldn't put it down. [1249]
>
> I saw it and first of all thought 'No way, it can't be done.' But, then I read it, I looked at the evidence and it was all there, it was well written and it was beautiful chemistry. And that's how I refereed it: this is beautiful, as is, with no corrections. There are things that have beauty, even in chemistry! [2138]
>
> In my case, in physics in terms of trying to understand it with the appropriate models and the description of it is all there in the paper – you have it all there in front of you. It's fantastic! [3145]

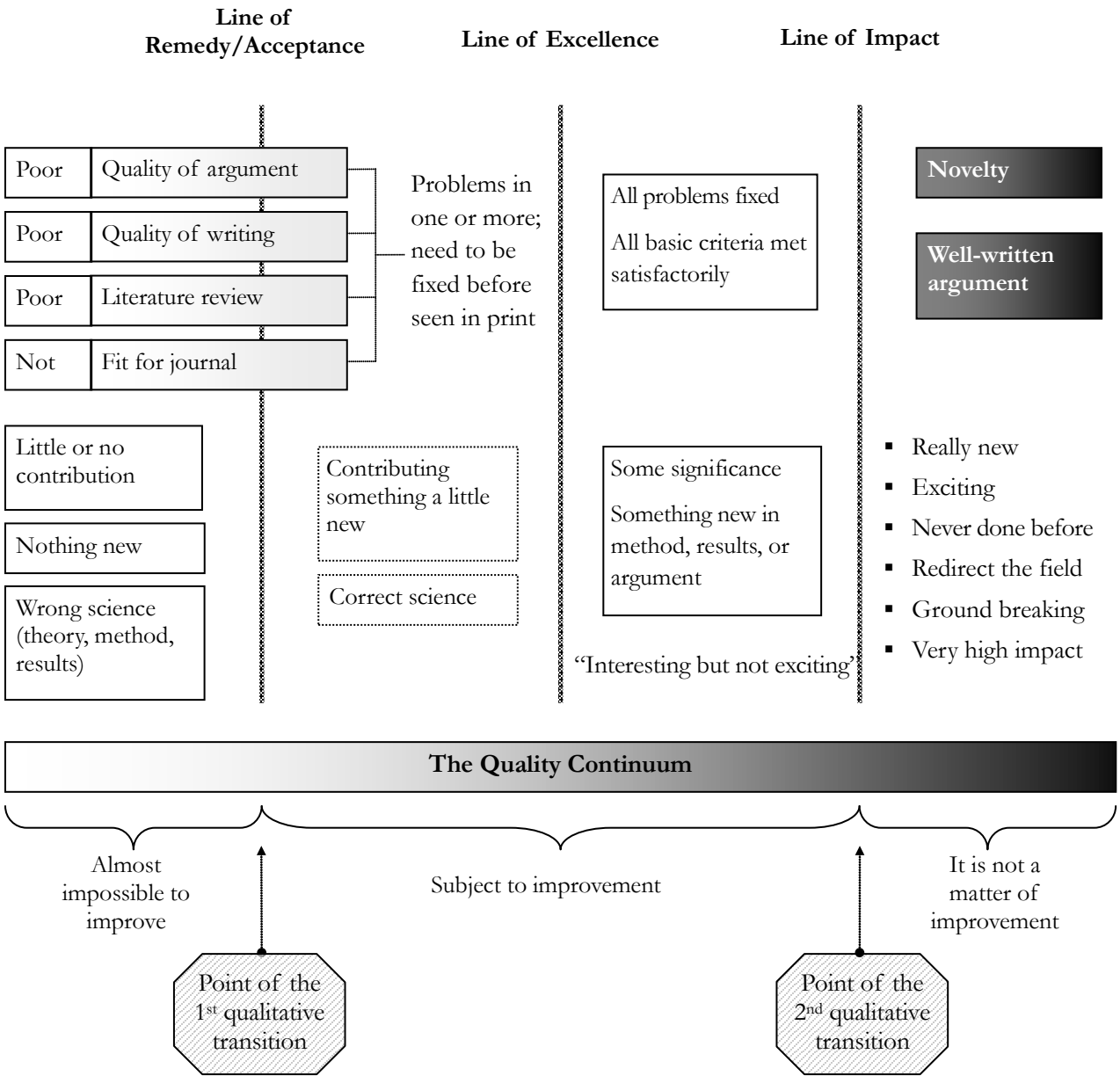
To sum up, to be recognized as one of the 'best', a paper needs to have 'an element of surprise' [3103] in what it has achieved. It also needs to articulate the contribution in an enjoyable, easy-to-read manner, which appears to be a reversal of the old saying: 'If you can't convince them, confuse them.' It is one that 'has made a significant contribution and it is going to have major influence on other people's work, coupled with that will come the quality of the writing and the presentation of the results'. It is not only 'the findings, but the whole body of what is presented will be of a very high quality for it to be in the best research paper category' [2215], that is, where reviewers draw the Line of Impact.

7.5 The 'Quality Continuum'

The responses indicated ways of stratifying papers of different quality. Reviewers' conceptions were consolidated into three lines in Figure 7.5. The *Line of Remedy* marks the threshold between outright rejection and marginal acceptance; the *Line of Excellence* marks the threshold between marginally acceptable quality and high

quality; and the *Line of Impact* points to one level further. Papers passing beyond this line are bound to show impact sooner or later, and milestone a major qualitative transition from good to extraordinarily high quality.

Figure 7.5 The Quality Continuum



As Figure 7.5 shows, a paper that is barely remediable, namely, positioned towards the left side of the *Line of Remedy* could be 'a replication of previous studies' which just 'located the study in a different cultural location or a different age group of people they may be working with' [1126]. The paper moves toward acceptance if:

> It contributes something that seems a little bit new, a little bit different compared to what is known about in relation to that small section of the field that they're writing about. It doesn't have to be something of world shattering importance or a major breakthrough. [1133]

Such a paper needs more work to move towards the right along the *Quality Continuum*; its authors need to for example 'make their literature review more substantial or they should refer to some other findings, or tighten up their argument in relation to how they collected their research data and so on' [1133]. A paper of this calibre may still be 'mundane… it's like 'I've read this before. It's just a different form of it.' It may say something new but doesn't do it in any special way, or any special form.' [1159]

A paper would be placed around the *Line of Excellence* if it 'has clear ticks on all of those [criteria] and is well written and well structured; it's clear and simple in its line of development and its coherency' [1210].A paper will be positioned around the *Line of Impact* if it not only offers 'something really fresh and new' [1159] and 'really of great interest in a broad context' [2215], but also

> satisfies all the criteria and it does that in a fresh and interesting and usually authoritative way; it appreciates across all the literature in the field; it has a strong understanding, there is a sense of direction coming from it, and it is extremely well written. [1159]

Such a paper 'has to not only be interesting and insightful in itself, but also provides things that are of use to other researchers in that field' [2173], and it won't be 'buried away' [3102]. It is laden with significant impact which leads reviewers (and certainly readers) to 'wonder how I can integrate that information into my work' [1218]. A metaphor used by a reviewer explained this sense of impact:

> It's a little bit like going to a movie. Some movies are really rubbish and you walk out and you forget about it in a few minutes. Whereas some movies leave a lasting impression on you and you remember it forever more. [1218]

Towards the far right end of the *Quality Continuum*, there are features denoting the 'best' papers. While many of the reviewers indicated that they would accept a paper as 'high quality' even if it was not that ground-breaking, they did expect to read something 'incredibly interesting' [1126], something that 'convincingly changed the way we think about a certain issue' [2153], something 'redirecting the field' [1146], and something they themselves 'had never thought previously' [1169]. Some of the reviewers used vague expression when describing the quality of the 'best' papers, and one actually asserted that it 'had to be vague' [3102]. But on the whole the stellar qualities of the best were unmistakable in the tone.

The distinction between qualities will be further explored in the next chapter.

7.6 Comparison between Journal Peer Review and ARC Grant Peer Review

The last two interview questions have touched the other major application of peer review, grant peer review, as adopted by the ARC. The respondents were invited to compare the two types of reviewing. The aim was to determine any yet unidentified features of journal peer review by virtue of comparison. A total of 33 respondents provided responses, including 19 from education and 14 from physics and chemistry.

The responses revealed that the similarities of the two types of reviewing were mainly in the consideration of quality, 'you're looking for the quality of the research' [2215], 'originality' [1126] [2166], 'importance' [2153] [3183], 'whether there is a potential interest' [2173], 'reference to the literature being made completely' [2153], and 'well written and clearly presented' [3183].

The considerations were not unlike those identified previously. Both types of reviewing required a demonstration of good 'academic position' [1202] or 'academic rigour' [1221], in the sense that 'you are at the cutting edge of your field, that you have a very strong knowledge base, that you have done research in the area, that you fully understand the concepts and that you have the capacity to carry out that research task.' [1221] Good 'academic position' or 'rigor' is a requirement for the competence of the author or grant applicant to be a qualified researcher.

All the respondents have noted some differences between journal and ARC grant peer review. The major differences between the two were procedural variations. Four major differences of this kind were identified. First, ARC assessors read research proposals; they are 'guessing' and 'it's a matter of choosing the ones that are most promising' [1218], while journal reviewers read 'reports on completed research' [1146]. This brought about a number of quality-related concerns among reviewers. An editor, reviewer, and ARC assessor-respondent offered some good examples of such concerns:

> Many proposed research methodologies in ARC projects were founded on the rocks of reality. The research methodology may be set up one way but turns out to be very difficult to do that way, so they do it another way. So it's a bit difficult to assess what's actually going to happen. While you are reviewing journal articles, you are looking for contributions to the field; but ARC is a bit more difficult because it's much more speculative, in that, people say what kind of contribution they hope to make, but in many fields the kind of evidence you turn up, the data you turn up, points you in a completely different direction. [1146]

Secondly, since the ARC reviews proposals, it 'has to assess the feasibility, and a big part of feasibility is the track record of the applicants' [2107] and whether a proposal 'is achievable by those people within the particular time frame' [1126]. Track record is 'treated in a very major way' [2153], 'counting for 30 or 40 percent of the weighting', whereas research approach counts 'only for 20 percent' [1221]. In addition, an ARC application has to 'be right on the money and with their National Priorities' and 'focused on getting it right and getting what they want to be researched' [1221]. None of these was deliberately considered in the reviewing for journals.

Thirdly, the ARC evaluation process involves 'a ranking process' in which assessors 'compare one person's argument of research design against another person's, whilst for journals 'reviewers evaluate the extent to which this particular writer contributes to the field, and not necessarily make a ranking that one person contributes more than somebody else.' [1126]

Finally, the ARC involves a higher level of competition than journals in general. Being the sole funding body for most researchers in Australia (the other national funder is the National Health and Medicine Research Council, which caters chiefly for medical and clinical research), the ARC applies a strict proportion of the application projects to be funded annually. As a result, 'if you don't get an ARC grant, you don't get your research' while 'if you get rejected in a journal, there are a hundred of other journals where you can try to get your article published in' [1185], and 'in some cases, the general interest and significance of a paper might compensate for some of its weaknesses' [1150], however, a single criticism in a grant review could be enough to 'guarantee that [the project] won't get funded.' [1249]

The respondents seemed to bring different expectations to the reviewing of projects and journal articles, which echoed the initial survey results. It is interesting to note that, when comparing ARC and journal reviewing, they became much more explicit about their commitment to provide constructive reviews than in the responses to the previous interview questions (see Chapter 6). For example, 'The purpose of your review is to suggest to the authors improvement that will bring the manuscript above the threshold of the quality at which it can be published' [1218].

The respondents were more satisfied with their experience as an author than as a grant applicant because they could 'get quite detailed feedback, either on how to improve' the paper or when being rejected, 'actually see the detailed way of being assessed why it's been rejected' [2129].

In contrast, 'the culture of the ARC today is not 'how do we fund these projects?', but 'how can we reject these projects from funding?' [1185] A few respondents attributed this to the limitation of available funds. Some noted that, when reviewing for the ARC, they were rather more hesitant to offer advice for improvement as they would when reviewing for journals. Reasons for this could be 'the ARC assessment is not about a finished piece of work', intervention is less appropriate, so 'changes are not the emphasis of the ARC assessment'; and 'grant assessors might have 20 applications to write comments on in a very short period of time' while journal reviewers 'deal with one review in isolation and can write in much more detail' [2189].

Another reason could be that, even if an assessor has identified an area that can be improved, 'there's a bit of tension in suggesting ways in which it might be made better without suggesting that the proposal is weak. In the ARC, because the competition is so fierce and the funds really need to go to the very best work, you cannot really make those sorts of allowances' [1150] that are possible in manuscript reviewing.

The respondents also recognised subtle considerations in weighting similar features differently in the two contexts. This quote explains the emphasis on argument and results for a paper and the emphasis on significant research questions for a proposal:

> The ARC assessment looks for arguments about the significance and the importance of the problem, and I weight it very heavily as a reader of it. When I review a manuscript, I would look for similar things, the importance of the problem, but is more about the finding of that problem. It may not address a problem that I see as important, but a very strong case has been made for the problem. I accept that as a paper, which is different to an ARC project. [1126]

This variation was confirmed by other responses:

> The ARC says, "would this be good research to undertake, and therefore to fund?" The [journal review] says, "what did this research tell us?" and "is this a good way presenting the findings through this article?" [1185]
>
> Research grant review is really concerned with evaluating the research question and its relation to the literature, the research methodology, whether it is appropriate; [while journal review is] more about the overall argument and the outcomes and the context in which it is written. [1169]

In broad terms, journal reviewing 'is probably more micro-managing'; 'it doesn't seem to put a lot of emphasis on how important the work is; it is normally assumed' [2189] and 'a lot of [journal articles] are philosophical and it doesn't have to go anywhere except to challenge people's notions and ideas', while for an ARC project, 'it's got to be, "Is it doable? And if it is doable, what effect is it going to have?" so [the ARC peer review] is much more pragmatic.' [1202]

7.7 Conclusions

Findings discussed in the chapter confirmed that the assessment of research in the two contexts involved the same range of considerations. It was also evident that, compared with grant assessment, when acting as journal reviewers, the respondents generally felt more comfortable to help authors to improve their work, and they were in a position to filter merits out of flaws in the submission, that is, the flaws could be compensated by the merits, and to identify flaws that were remediable. The differences highlighted the collegial functions of and the stewardship involved in journal peer review, as discussed in Chapter 6.

Reviewers drew on a range of features to stratify papers by recommending some to be rejected while others to receive a major or minor revision. The broad groups of the features covered the same ground as criteria identified in previous studies, journal policies or editorials. Yet this phase of study confirmed that reviewers placed emphases on different features when making different recommendations.

Drawing on what the reviewers say, the process of evaluating a paper can be best described as a process of weighing up remediable problems against irremediable problems and this brings into play what can be envisaged as a quality continuum with two points of transition, the line of remedy, and the line of impact. In between these two lines, there is an internal dialogue about how to judge the relative impact of the flaws and strengths that shapes the reviewers' comments for authors to address.

CHAPTER 8: LANGUAGE OF REVIEWING: ANALYSIS OF REVIEW REPORTS

8.0 Introduction

In the interview the reviewers explained a range of subtle considerations in reviewing that now can be extended to the analysis of their reviews in this chapter. The researcher has identified a small number of studies that drew on reviews as a primary data source and some of these used content analysis. Bakanic et al. (1989) found that reviewers of a sociology journal were more likely to praise papers recommended for publication; and in general the positive comments focused more on style and the negative comments more on theory and statistical analysis. Fiske and Fogg (1990) focused on comments of weaknesses of 153 papers submitted to American Psychological Association journals and found the most frequently noted weaknesses were in the interpretations and conclusions and the presentation of the conceptual work, followed by criticisms about results, procedures and design. They revealed that reviewers saw as many problems in the presentation and description as in the actual research activities (p. 767). Similarly, Gosden (2003) analyzed 40 reviews of 21 papers submitted to a physics journal and revealed that reviewers focused more on the quality of writing and style than on method, analysis or theory, and were more straightforward and specific in their suggestions for the author if they favoured the paper.

In addition, two studies on the same journal revealed two reviewer styles, critics and coaches, with the latter tending to provide more constructive reviews which were more likely to lead to authors getting published (Beyer et al., 1995; Cummings et al., 1985). Another study identified no clear relationship between the tone, depth, or content of reviews and characteristics of reviewers, and the tone of the reviews did not necessarily predict the direction of the recommendation (Murphy & Utts, 1994). Some studies analysed reviews to test for the level of reviewer agreement and potential bias (Epstein, 1990, 2004; Gilliland & Cortina, 1997).

This phase was undertaken after the interviews were completed to gain some insight into the actual activity, especially with respect to emphasis on features of quality as well as to use another source to elicit further depth of understandings on judgments about quality. The researcher focused on evaluation language and how reviewers communicated their considerations about the quality of the manuscript to authors, primarily to determine if the content and language of reviewing reflected the considerations and the concerns raised by the participants in the interviews.

Unlike many other studies where editors had access to all reviews of interest, the researcher was faced with a number of restrictions on access. Because the choice of reviews was totally at the reviewers' discretion, the researcher has no direct control over the selection and therefore the type, thrust and quality of the reviews. In accordance with ethics requirements, the reviews were de-identified by the reviewers before they sent them, and so the researcher did not have access to information such as the journal for which a review was prepared or the final decision on the paper. Finally, because there could be no control over what reviewers would have received to review, the content of the reviews reflects variation in the quality of the papers.

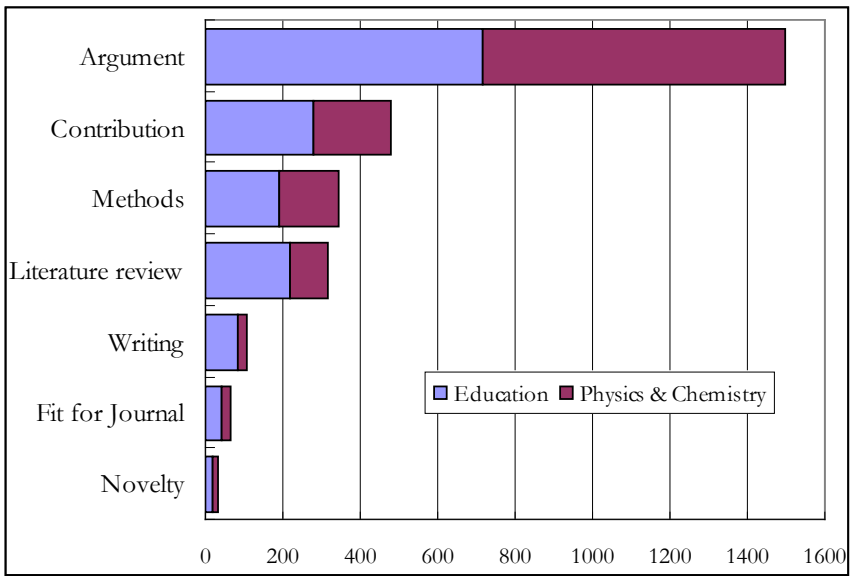
A total of 61 reviews were provided by 28 of the interviewees, 28 from education and 33 from physics and chemistry. The reviews were given an ID. For example, respondent 1150 provided two reviews – they were coded as 1150-1 and 1150-2; respondent 1159 provided one and it was coded as 1159-1. Then the text was coded in QSR-N6 by line. The 61 reviews consisted of 3165 lines in total, and varied in length, ranging from 8 to 170 lines, with an average of 53 lines.

Drawing on the literature and the analysis of the interviews for guidance, the text units were coded into navigational units of the seven pre-determined categories. No additional feature of quality was identified. Distribution of the text units in each category is illustrated in Figure 8.1, divided by discipline.

More than half of the text in the reviews was devoted to *argument* (52.8%); this was followed by *contribution* (16.7%), *methods* (12.1%), *literature review* (11.1%), *writing* (3.8%), *fit for journal* (2.3%), and *novelty* (1.1%). As 'novelty' could refer to new angle

of comprehending the literature, new methods, or new way of interpreting data, this category overlapped with the others.

Figure 8.1 Count of Text Units by Category



In addition to the proportions of text, each instance of a category was counted and assigned a 'tone', namely, whether the comment was predominantly positive or negative. 'Negative' indicates the need for further work or revision, positive indicates no need for this. If a review contained a combination of comments on a feature such as argument, which was rare, the dominant tone was noted. For example, a reviewer [1194-2] started one section: 'I believe this paper is worthy of publication in this journal' but continued, 'However, I would suggest some changes to…' and goes on to provide 36 lines of suggestions for revision, so this was recorded as 'negative'.

Tables 8.1-8.7 in the following sections present the count and the ratio of positive to negative comments in each category. Table 8.8 assembles all the relevant statistics together for comparison.

8.1 Comments on Argument

The interview data revealed that 'argument' was the most emphasised feature in the evaluation. The respondents drew on 'argument' and 'novelty' to differentiate high quality from marginally acceptable quality. In the reviews they prepared, 'argument'

again received the most frequent and intensive attention; comments about argument were found in 53 of the reviews (87% of 61), which contributed to 52.8 percent of the total text units. Of these, 23 were from education and 30 were from physics and chemistry (Table 8.1).

Table 8.1 Number of Reviews Containing Comments on Argument

Feature	Education	Phy&Chem	Total
Quality of argument	23 (3:20)	30 (1:29)	53 (4:49)

Comments on argument were primarily oriented to revisions, including explanation of problems, and suggestions for improvement. The four positive comments were all very short in length. The analysis of the reviews coincided with the interview findings in that the reviewers looked for 'strength' in the argument, including its *clarity* and *consistency*. Two quotes to illustrate this:

> *Clarity* – explanation of key terms such as 'scaffolding', 'social scaffolding' and 'analytic scaffolding' was very clear, and the examples provided to support these definitions add to understanding. [1224-2]
>
> *Consistency* – the study is well documented with conclusions both justified and adequately summarized. [1249-1]

The lack of clarity in the argument attracted most discussion in the reviews. The reviewers identified different instances of the lack of clarity. Four requirements for authors were most noted, including to: present more details to explain their points; organize the discussion in a better, logical order; support points with more critical thinking and analysis; and avoid overstatement and over-generalization. Most of the reviewers also offered suggestions for improvement, often by putting forward questions for authors to address.

'Insufficient detail provided' was the most noted deficiency related to the clarity of argument. Reviewers' criticisms included: the information provided was not enough for an appropriate understanding of the research, such as how the research questions were determined and 'what further questions remained to be answered' [1224-2];

how the theoretical framework was developed; why a term was chosen for a conceptualization; how the participants were selected; and how the experiment was conducted or its results measured. Several reviewers asked authors to 'publish the survey questions' [1126-2] or the procedures of the experiment.

In order to address this problem, reviewers typically required authors to provide 'further elaboration' [1126-1] or 'justification' [1210-1], 'more key examples' [1147-1], 'formal analysis of the phenomena' [1194-1], a 'description of the diagrams' [1242-2], 'a transcript number' for each quote [1150-1], and a better discussion of the 'significance' of the study [3165-1].

Many reviewers identified very specific areas in the paper where more details were needed. An example of this was, 'Table 1: it would be informative to report % residues in the different regions of…' [2107-2]. Some reviewers indicated that they felt 'confused' [3110-1], 'disappointed' [1126-1], or were 'struggling to discover anything new' [1126-2] when such details were missing.

'Lack of logical flow' was the second major concern in relation to the clarity of argument. Comments included for example,

> Some connections have been 'forced' and these hamper the establishment of a logical progression. Straw men are set up and then demolished in support of an 'argument'. [1147-3]

Some picked up gaps and contradictions between points, such as,

> They say it is impossible to use the [data] with their method, but their Paper II does this! [3183-1]
> The notation for column vectors introduced on the previous page has not been used. [3110-1]
> Instead of analyzing this rather important point, the authors discuss [a criterion] not previously mentioned [1150-2]
> It would have more impact if it were more clearly structured, perhaps dividing it into sections each with a rationale that derives from the previous discussion. [1249-1]

Some reviewers identified cases where the connection between core components of the research (for example, the theory and the instrument) was 'vague', without being built on 'sufficient detail' [1150-1]. Some reviewers tried to help authors to improve the flow by providing specific suggestions such as:

> Move the final sentence of this paragraph to after example (10) [1194-1]
>
> Introduce the descriptor 'periclase'… earlier in the paper [3102-2]
>
> Discussion of the effect of the… belongs to the Results and Discussion section. [3180-1]

'Lack of critical thinking and analysis' was the next most cited concern regarding the clarity of argument. The reviewers required authors to make their points stronger and more sustainable, so as to give the paper 'an edge of contemporary relevance that will make more people want to read it' [1244-1]. These three quotes illustrate the issue:

> The focus on the rating scale means is not very helpful, both because of the data itself and because of the lack of critical analysis of it. [1150-2]
>
> The potential strength of the paper lies in reporting what is NOT happening. However, the discussion and conclusions reflect a very descriptive analysis of the data and not a deep reflection on what might be useful outcomes of the data. [1126-2]
>
> Simply observing such fluctuations without exploring their connection with ion permeation… appears to me as a wasted effort. [3180-2]

Lastly, in a few cases, the reviewers detected overstatement and over-generalization which made the argument 'doubtful' [3180-2]. They required authors to amend or remove them, for example, '… is perhaps too strong a statement' [2189-1] and suggested alternative forms of expression. Another example was: 'While [X] are certainly an excellent way to eliminate daily aliases, I think it is too strong to say that they 'must be used'. I prefer 'can be used', since a ground-based network similar to [Y] would also do very well.' [3183-1]

Consistency was the other major expectation of the argument. As the reviews show, the consistency, and reliability, of the argument is about how well the conclusion is supported by the evidence and how well this link is presented in the paper. Reviewers became very critical if they identified a lack of (adequate) evidence from which a conclusion was drawn. The word 'convinced' or similar was common in the reviews:

> I'm not sure how such a conclusion can be arrived at. Where's the evidence? [1147-1]
>
> I'm not convinced. The rationale for such an endeavour is non-existent. [1147-3]
>
> I am not entirely convinced, on the basis of the evidence presented. [1194-1]
>
> The discussion describes aspects of the program in rather glowing terms, without providing evidence for the claims. [1150-2]

Some reviewers required authors to present a better conclusion from the evidence, even though in some this was not going to lead to acceptance, for example they may also have noted 'that they cannot answer this question based on present data' [2107-1]. Others sought better integration of the data. For example:

> More discussion of the issues you raise in this paragraph seems important in framing the paper and the conclusion would be stronger if it returned to this and the 'hollowness' of naturalistic research that you refer to, and the consequent… that you found. [1255-1]

To sum up, in the reviews collected for this study, reviewers looked for clarity, logic, sustainability, adequacy, and consistency in argument. Some of the most common deficiencies that authors would make included: failing to provide sufficient detail for key components of the research, and drawing ill-supported claims and conclusions. As a characteristic of features related to remediable problems, comments about argument were substantiated with extensive suggestions for improvement.

8.2 Comments on the Review of Literature

Quality of the 'literature review' was discussed in 36 or 59 percent of the 61 reviews; the comments contributed to 11.1 percent of the total review text. As Table 8.2 shows, the comments were mostly negative; 21 reviews were from education and 15 were from physics and chemistry.

Table 8.2 Number of Reviews Containing Comments on Literature Review

Feature	Education	Phy&Chem	Total
Literature review	21 (4:17)	15 (0:15)	36 (4:32)

The four positive comments were all from education, and according to the pattern already established above, were all very brief and generic. A sound review of the literature in education was described as one that is 'relevant and recent' [1150-2], 'contains current thinking in the field… synthesizes the main idea' [1224-2], 'shows evidence of good background research' [1226-1], 'makes innovative use of philosophers/theorists who have not been taken up much in education' [1255-1], and 'well written' [1224-2] [1255-1].

The negative comments were focused on three major problems: the neglect of useful literature, the inappropriate inclusion of literature, and the insufficient critical analysis of the literature.

The most common problem in terms of the literature review was the 'neglect of useful studies', where 'authors need to give better credit to previously published work' [2215-5]. The reviewers either pointed out places in the paper where more reference to the literature was needed, or noted studies that authors should have considered. Those from physics and chemistry were more likely to provide a list of references for authors to consider. Some required authors to build better links between the literature and their research by '[detailing] the construction of the survey questions in the light of the theorization of the field' [1150-1] and 'revealing where, conceptually, that problem came from' [1244-1].

Some reviews noted the problem of making uninformed claims, for example, 'the problem is simply described. There are concepts that are not problematised about team-based assessment from the relevant literature' [1259-1] and 'statements like [X] are misleading if not included in a historical perspective' [2215-5]. Authors may have introduced an important theory or concept, but failed to 'define' or 'explore in the detail it deserves' [1226-2] by investigating the literature thoroughly. One reviewer in education noted it is a problem if the author 'introduces their notion' 'without satisfactorily clarifying how it differs or [is] similar to the earlier theories' so 'the argument is somewhat confounded'. [1126-1]

In addition, the practice of self-citation was criticized, 'little of the literature outside the author's own previous work and that of colleagues is cited let alone drawn upon for comparison' [1126-1]; it 'indicates that other workers in the field are likely not given sufficient credit.' [2107-1]

The second common problem was the 'lack of critical analysis of the literature'. It was noted in eight reviews, all from education. Some reviewers described such a review as 'shallow' [1126-1], 'unsophisticated' [1150-1], 'over generalized and, in some sections, relies excessively on the work of others' [1255-2]. As one wrote,

> Rather than presenting a well reasoned argument for something, it seems to be pursuing more of a 'walk through the literature'; everything seems to have been thrown in, with much of it not supporting anything specific. [1147-3]

The third problem with literature review was the 'inappropriate inclusion of literature', noted in six reviews, five of which were from education. These reviews singled out references cited that did not '[add] much to this paper' [1126-1], that were 'dated' [1159-1], 'not really necessary' [1221-1], or not 'meaningful' [2166-1].

8.3 Comments on Writing

Fifty reviews commented on writing, including layout and referencing style. Some reviews were extended by lines or even pages of dot points to correct spelling, punctuation, and Basic English (for example, the use of 'the'). Other reviewers noted

the need to 'correct all this on the document' [1242-2], which indicated that the total amount of corrections was much more extensive than appeared in the reviews. The remaining comments on writing were found in 23 reviews, or 3.8% of the total text. As Table 8.3 shows, negative comments were twice as frequent as positive ones, and reviews from education appeared to be more negative in general.

Table 8.3 Number of Reviews Containing Comments on Writing

Feature	Education	Phy&Chem	Total
Writing	14 (4:10)	9 (4:5)	23 (8:15)

The eight positive comments were all very short; yet they offered an immediate idea of what 'good writing' should look like. A well-written paper is 'clear' [1150-2] [1249-3], 'accessible, highly readable' [1249-1], 'beautifully crafted and easy to read' [3102-1], 'concise but still comprehensive' [3183-1] [3183-2], and '[doing] very well in using Australian language to provide the bulk of the empirical evidence' [1194-1].

In contrast, the negative comments on writing were much longer and all very critical, especially those about use of Basic English. In some cases the comments conveyed a sense of the irritation of the reviewers. For example, a reviewer wrote that:

> The author should realize they have a responsibility to their colleagues who serve as referees to submit manuscripts in which the grammar is reasonable… The grammar needs considerable polishing. [3102-2]

Another asserted that 'There were many technical errors, for example, punctuation, referencing which leads me to conclude that this was not a scholarly paper' [1159-1]. And a third, after reading a paper submitted by non-native English speaking authors in which 'presentation was extremely poor with unreasonable numbers of basic mistakes in English', wrote that,

> I corrected them all to begin with, but gave up about half way through as there were so many. I find it amazing that authors would send a paper overseas without any proofreading… Tackling all this and trying to make sense of it all, was a complete waste of my time.' [1242-1]

Other negative comments included criticisms towards the use of jargon, which could involve unnecessary expressions such as "'evidentiary warrants" – what is wrong with "evidence"' [1126-1] and problems in the 'expression of dates' which is peculiar 'for an international audience' [1126-1]; and the use of conversational style, such as 'got the spotlight on the right place' [1147-1] and 'direct speech statements' [1126-2].

8.4 Comments on the Fit of the Manuscript for the Journal

'Fit for journal' refers to the suitability of a manuscript to the focus and audience of a journal or its conformity to the submission format required by the journal. This feature was discussed in 11 reviews; six from education and five from physics and chemistry. The relevant comments made up 2.3 percent of the total text units.

Table 8.4 Number of Reviews Containing Comments on Fit for Journal

Feature	Education	Phy&Chem	Total
Fit for Journal	6 (0:6)	5 (2:3)	11 (2:9)

Two reviews, both from physics and chemistry, commented positively as 'The work is clearly within the scope of this journal.' [2107-2] The other nine wrote negatively as, 'I am not convinced that this journal is really the journal for the paper'. [1126-1]

The reviewers indicated in the interviews that they would reject a paper outright if it was outside the scope of a journal. However, the negative review comments revealed a more generous intention in practice. Instead of suggesting rejection, five reviewers required the authors to improve certain aspects of the paper to make it more suitable for the journal, for example,

> I would suggest that some additional discussion of the significance of the results be given so that the paper is more suited to the main journal. [3165-1]
>
> If more examples of events concerning science learning were included in the paper, it would be far more interesting to readers of the journal. [1126-1]

The other four reviewers did recommend rejection, but all suggested alternative journals that they thought were more suitable for the paper to submit, commenting, 'The work is very specialized towards imaging and spectroscopy and could gain much better impact in a more specialized journal in those areas.' [2138-2]

The foregoing four sections discussed comments in the reviews on the four features related to remediable problems (Group A). They shared two common characteristics: positive comments were always very short, and the criticisms were always followed by suggestions for improvement, often in the form of questions for the authors to address. Comments on the other group (Group B) of quality features, including contribution, research method and novelty, are analyzed in the following section.

8.5 Comments on the Contribution of the Research

Evaluation of 'contribution' was located in 48 reviews (79% of 61), 24 from education and 24 from physics and chemistry; the comments made up 16.7 percent of the total text units. While the reviewers did not provide much explanation about contribution in their responses to the interviews; this was more in evidence in their reviews.

Table 8.5 Reviews Containing Comments on the Four Types of Contribution

Contribution	Education	Phy&Chem	Total
Topic selection	12 (8:4)	1 (1:0)	13 (9:4)
Relevance to practice	11 (3:8)	1 (1:0)	12 (4:8)
Advance of knowledge through findings	8 (3:5)	14 (9:5)	22 (12:10)
Advance of knowledge through approach	2 (2:0)	8 (4:4)	10 (6:4)
Total	**33 (16:17)**	**24 (15:9)**	**57 (31:26)**

The analysis revealed four types of contribution, including: whether or not a paper 'addresses an important topic' [1150-1] or 'raises an important issue' [1226-1], whether the findings advance the knowledge 'beyond what is already well known' [1126-1], and whether the paper has meaningful relevance 'to the extent that readers from elsewhere will engage with' it [1210-2]. The reviews were coded for four forms of contribution – topic choice, relevance to practice, and advance of knowledge by providing important findings or by adopting a certain approach or method.

Table 8.5 presents the distribution of reviews by theme and discipline. Many reviews discussed more than one type of contribution and they were allocated to multiple groups; as a result, the total count exceeded the number of reviews in this category. As it shows, there was a marked disciplinary variation in reviewers' emphasis: education reviewers prioritized contribution in 'topic choice' and 'relevance to practice', while physics and chemistry reviewers prioritized contribution in the form of 'adopting certain research approach' and 'providing important findings'.

Contribution regarding 'topic choice' was discussed in 13 reviews, mostly from education. The negative comments were mainly about the lack of 'direction' given in a paper, in the sense of being 'not sure what the purpose of the article is' [1147-3]; they asked 'what is the big question that this paper addresses?' [1249-2]; and they suggested that 'the authors might reconsider exactly what their focus is in the paper. At the moment focus is spread too broadly.' [1255-2]

The positive comments on 'topic choice' used two words: 'interesting' and 'important'; no one offered further explanation. When reviewers found the topic interesting or important, they pointed it out immediately, even though they may have remained critical about the rest of the paper. Two quotes illustrate this:

> The article does address an important issue in early childhood regarding motivation. However, greater emphasis could be placed on… [1221-1]
>
> The [topic] is an interesting concept worthy of investigation. This paper, however, does not… demonstrate clearly enough… [1226-2]

Contribution as being 'relevance to practice' was recognised mostly by education reviewers, and the ratio of negative to positive comments was the highest of the four. Relevant criticism includes:

> This case study does [no] more than say we have created a course that we really think is excellent. [1159-1]
>
> The context is very parochial – [name of a state]. It assumes that what is going on in [the state] is significant for the rest of the country/world… It also assumes that everyone will know and understand curriculum developments in [that state]. [1249-2]

Four reviews contained positive comments on 'relevance to practice'. Examples include: 'an excellent paper that addresses an issue that is both a strong theme in the literature and is a pressing concern for…' [1249-1]; and 'the data will allow researchers in other countries to compare and contrast trends in…' [2189-1]

Comments on 'providing important findings' were found in 22 reviews, 8 from education and 14 from physics and chemistry; the latter had a much higher positive to negative ratio. Studies that received praise for findings had either added new knowledge to what was known or challenged old beliefs with well-supported evidence.

In terms of 'adding new knowledge', the reviewers made comments like:

> This paper describes a useful extension to the chemistry of [X]. [2129-2]
>
> The authors do obtain a substantial result that could lead to a deeper understanding of [Y] and positive definite functions in [Z] [3110-1]

As for 'challenging old beliefs', there were comments such as,

> Many writers have assumed that… This paper demonstrates that this is not indeed the case. [1194-1]
>
> The results of this study highlight that our understanding of… must progress from the simple assumption that… [2189-2]

Some other reviewers just wrote the paper 'should make a very useful contribution to the literature' [2215-2] without explaining why.

Negative comments on findings often recognised authors' failure to add to new knowledge or challenge old beliefs. Examples include:

> I had considerable difficulty answering the 'so what' question. The findings of the study replicate common knowledge about… [1126-2]
>
> A protein of unknown function [was described]. However, preliminary functional assays on a few possible substrates have not shown any activity, and the substrates for this hypothetical enzyme remain unknown. [2107-2]

The reviewers were also not content with new yet insignificant findings, noting that: 'The authors admit that they cannot actually answer this question based on present data; but even if they could, would this be an important conclusion?' [2107-1]

Contribution concerning 'research approach' was discussed in 10 reviews and much more often by physics and chemistry reviewers. The reviewers praised a paper if it described an 'excellent application' [2138-2] or 'a detailed study' [3180-1], or it adopted 'a new way' [3183-1] of investigation. The following quote offers insight into one such new way:

> This paper is a radical and highly original theoretical and methodological invention… [It] makes innovative use of philosophers/theorists who have not been taken up much in education. [1255-1]

On the contrary, four reviews, all from physics/chemistry, criticized the approach as being not rigorous enough to make a contribution:

> The scattered experiments done do not form a coherent whole nor allow significant conclusions to be drawn. [2166-3]
>
> The proofs are elementary, use standard techniques and contain errors. [3138-1]
>
> Simply carrying out [a physical application] without making any effort to link them to experimental observations is not something I would recommend for publication. [3180-2]

In sum, the comments on contribution reflect the two common characteristics of Group B features: the judgment on contribution is constructed in a dichotomous manner: a paper makes a contribution or it does not; and in the comments there is an absence of justification for the judgment or suggestions for improvement. Some reviewers did suggest that authors try to improve their argument so as to highlight the contribution, but as for contribution *per se*, there was no comment on improvement. Once again, it is important to draw attention here that this is not a representative, but a volunteered group of reviews.

8.6 Comments on Research Design and Methods

'Research design and methods' was another Group B feature. Comments on this feature were found in 28 reviews (46% of 61), split evenly between education and physics and chemistry; they contributed to 12.1 percent of the total text units.

Table 8.6 Number of Reviews Containing Comments on Research Design & Methods

Feature	Education	Phy&Chem	Total
Research design & methods	14 (2:12)	14 (2:12)	28 (4:24)

Seven reviews made judgments directly on methods. Negative comments judged the methods as to be inappropriate or dubious including:

> The fact that only nine teachers responded to the survey makes the close attention to the rating scales questionable. [1150-2]
>
> Author of this paper seems [to] take an entirely uncritical or unreflective stance in relation to the views of these teachers. [1249-2]
>
> I don't believe that it is possible to inject aggregates of this kind gently enough so that they are not damaged. [2166-1]
>
> It is not very good practice to calibrate on a single standard' [2166-2]

Positive comments on method included, 'The studies have been done very carefully and the quality of the data is very good' [2215-1] and 'The experimental work is competently performed' [2215-5].

The other reviews, including two positive and 21 negative ones, did not offer a direct evaluation on methods. Instead, they commented on the description of the methods, such as whether it was 'well-substantiated' and the 'strengths and limitations of research approach are explored' [1226-1].

In other words, these comments focused more on how well the research was reported rather than how well it was conducted. An explanation for this was, 'As a reader I need to know much more about the conduct of the study, who the participants were and why I should place any credence in what they said'; if the

information was lacking, it could put 'the credibility of the research' in question 'as there is no basis for judging whether the research is credible or not.' [1249-2]

The reviews raised questions about all aspects of the reporting of the research that could be impaired by insufficient description. In terms of the 'rationalization of the research', several reviewers criticized the lack of discussion 'on why the survey questions were chosen' [1150-1]. In terms of the 'procedures of investigation or experiment', the reviewers generally required authors to provide more specifics.

In some papers some very basic information was 'barely discussed' [1255-2], including 'research subjects and site' [1255-1], 'when and to whom questionnaires were administered [and] what was the structure of observations' [1210-1], 'How were [the subjects] identified? What criteria were used?' [1224-2], what was 'the internal standard used for determining the cell parameters of the starting material' [3102-2], and 'What was the spot size for the UV laser?' [3138-3]

With regard to the 'presentation of data', reviewers required more detail about, for example, 'the range or standard deviation of responses' [1150-1], the 'numbers of respondents' [1210-1], 'the diagram' [1242-2] and 'the characterization of the resulting complexes' [2129-2]. According to the reviews, the lack of sufficient description of the research would disqualify a paper in many ways, among these were the difficulty 'to see how these pieces of work relate to the principles' of a research tradition [1242-1]; the potential of attracting 'a claim of bias [about] representativeness of the quoted opinions' [1249-1]; the absence of 'guidance to others who might want to apply a similar approach' [1150-2]; and, a hindrance to making an impact as 'it is not clear how this paper adds methodologically to the field' [1255-2].

Sound methods were one of the key features that would take a paper across the 'Line of Remedy'. Reviewers rarely offered suggestions on how to correct problems in the methods. They concentrated more on the description of the methods, pointed out the unsaid or questioned the unclear. In general, there were more, and often more specific, questions being noted in the reviews from physics and chemistry.

8.7 Comments on Novelty

The third Group B feature is novelty. Seven reviews discussed this feature, three from education and four from physics. The comments contributed to 1.1 percent of the total text (Table 8.7). Five reviews discussed novelty in research methods, including 'new ways of researching and understanding [X in] ethnographic research' [1255-1], and 'a new way of analyzing stellar oscillation data' [3183-1]. The other two identified novelty in findings, simply stating that 'This paper presents new results' [3165-1] or 'there is very little new or worthwhile in the paper' [3180-2]. Three commented positively and the other four were negative.

Table 8.7 Number of Reviews Containing Comments on Novelty

Feature	Education	Phy&Chem	Total
Novelty of methods	3 (1:2)	2 (1:1)	5 (2:3)
Novelty of findings	0	2 (1:1)	2 (1:1)

The comments on novelty, although scarce and brief, revealed at least four characteristics of this feature. First, novelty is a rare merit to achieve. This was implied in the interview data and was confirmed by the reviews – only two reviews judged a paper as presenting something of novelty.

Second, as a Group B feature, none of the reviews provided any suggestion to 'improve' the novelty of the work. Third, negative comments on novelty relate closely to outright rejection or, rather, reviewers often use 'lack of novelty' as the reason for outright rejection – all the five reviews commenting negatively had recommended rejection, and the recommendation immediately followed the comments on the lack of novelty.

Lastly, with respect to the manner in which novelty was discussed in the reviews, the reviewers normally made a straightforward statement and cited the word 'new' or 'novelty' directly, without explaining much about what their expectation of novelty would be; as these two quotes illustrated, 'So what is new? [The method] is relatively

commonplace' [1126-1], and 'There is very little new or worthwhile in the paper' [3180-2]. The very little explanation of 'novelty' and what it means in the reviews suggests that while novelty can be recognized by the reviewers by drawing on their knowledge, they do not choose to articulate the knowledge in-depth.

All the foregoing review and interview data were brought together for comparison. Section 8.8 below discusses comparison within the review data; Section 8.9 presents the findings of the comparison between the interviews and the review data.

8.8 Disciplinary Comparison for Remediable and Irremediable Problems

Tables 8.8 and 8.9 consist of summary statistics that facilitate a broad-brush comparison of emphasis in the reviews. The first table presents the number and percentage of reviews that fell in each category and the ratio of positive to negative comments by discipline. The second table presents the count of text units in each category. So, the two tables cover frequency and length. The columns are horizontally divided into Group A and Group B features. Built on these tables, a comparison by discipline and across categories was undertaken.

Table 8.8 Number and Percentage of Reviews in Each Category Divided by Discipline

Features		**Number of Reviews (Ratio of Positive to Negative Comments)**						
		Education			**Phy&Chem**			**Total** (Ratio)
		N=28	%	Ratio	N=33	%	Ratio	
Group A	Argument	**23**	*35.9*	0.15	**30**	*50.8*	0.03	53 (0.08)
	Literature review	**21**	*32.8*	0.24	**15**	*25.4*	(0:15)	36 (0.13)
	Writing	**14**	*21.9*	0.40	**9**	*15.3*	0.80	23 (0.53)
	Fit for Journal	**6**	*9.4*	(0:6)	**5**	*8.5*	0.67	11 (0.22)
	Total	**64**	***100.0***	**0.21**	**59**	***100.0***	**0.16**	**123 (0.18)**
Group B	Contribution	**24**	*58.5*	0.94	**24**	*57.1*	1.67	48 (1.19)
	Methods	**14**	*34.1*	0.17	**14**	*33.3*	0.17	28 (0.17)
	Novelty	**3**	*7.3*	0.50	**4**	*9.5*	1.00	7 (0.75)
	Total	**41**	***100.0***	**0.51**	**42**	***100.0***	**0.83**	**83 (0.70)**

Some clear patterns of difference in the comments were revealed by the split of Group A and Group B features. For example, comments on remediable problems were not too different from those on irremediable problems in number, but were double the latter in length. This can be explained by the reviewers' tendency to offer extensive explanations and suggestions when they dealt with remediable problems, while such detail was largely absent from comments on irremediable problems.

Table 8.9 Count of Text Units in Each Category Divided by Discipline

Features		**Number of Text Units**			
		Education	**Phy&Chem**	**Total=2,839**	
				N	*%*
Group A	Argument	714	784	1,498	*52.8*
	Literature review	217	98	315	*11.1*
	Writing	84	24	108	*3.8*
	Fit for Journal	40	24	64	*2.3*
	Total	**1,055**	**930**	**1,985**	-
Group B	Contribution	279	200	479	*16.7*
	Methods	191	152	343	*12.1*
	Novelty	18	14	32	*1.1*
	Total	**488**	**366**	**854**	-

The ratios of positive comments to negative comments also show a difference. For Group A features, the reviewers tended to focus much more on problems than on merits. When a paper demonstrated high quality in any of the features, the reviewers only acknowledged it briefly, while if a problem was detected, they would always explain in detail why it was a problem and how it could be addressed.

By contrast, for quality features in Group B, although negative comments still outnumbered positive ones, the counts were similar. If a paper was judged as having made a significant contribution, for example, more often than not the reviewer would explain why the contribution was significant.

Compared to the distinctions between Groups A and B, the disciplinary variations were much more tentative. The most obvious ones included that the education reviewers more often commented on 'literature review' and 'writing' (Figure 8.1), and their comments were much longer, than the physics and chemistry reviewers.

When commenting on 'contribution', their emphases were different, too, with the former focusing more on 'topic choice' and 'relevance to practice' and the latter more on 'findings' and 'research approaches'. Other than that, the reviewers from the two disciplines demonstrated rather more similarities than differences in the interpretation of quality.

8.9 Conclusion: Comparison between Interview and Review Findings

The analyses of interviews and reviews provide two distinct, interrelated angles to address the questions 'What is quality?' and 'How are judgments and distinctions drawn?' in the context of reviewing. The interviews revealed reviewers' conceptions of the key features that a research paper should demonstrate, and the analysis of the reviews drilled into their interpretations of the conceptions in practice, as well as suggested ways about appropriate interpretation of the language of reviewing.

The review data affirmed the divisions between Groups A and B features identified in the interviews in three major ways. First, the reviewers approached Groups A and B features in their reviews with different emphasis. For Group A features, they commented very briefly on the merit while focused much more on the problem, by asking questions and providing suggestions. Such comments might appear to be harsh and difficult to reconcile on the part of the author, especially in the case of a lengthy list of corrections for the argument or presentation.

When commenting on Group B features, the reviewers would always acknowledge the merit in some length, such as explaining why they thought a paper had made a significant contribution. However, they were less likely to offer detailed explanation of, or suggestions for, dealing with the problem. The interview data revealed that the lack of explanation was due to the nature of these features. Nevertheless, it could

possibly cause confusion and attract complaints from authors such as judgment without justification or unfair judgment.

Second, the review data showed that the quality features could be further stratified into multiple 'quality specifics'. For Group A features, such specifics had already been clearly identified in the interviews. The review data substantiated them with more detail, such as the range of meanings attached to the 'clarity' of argument. This suggested that the reviewers established their conceptualization of these features in a generic and stable manner. The conceptualization and its specifics did not vary much by discipline. By contrast, when Group B features were discussed in the interviews, the reviewers provided only broad and brief explanations. In the reviews, discussions of them varied widely in focus and implied instant interpretation of the reviewers.

Third, the review data provided further detail for each of the quality specifics. It would appear that, aside from variations between Groups A and B, there were also variations between the specifics within the same feature. A summary of the specifics is presented in Appendix 15. The following sections discuss the variations by linking the reviewers' responses to the interviews with their comments in the reviews.

8.9.1 Group A: Features related to remediable problems

Argument is the most heavily weighted feature in the evaluation of a paper. The strength of argument lies in its clarity, logic, consistency and substance. Whether the author has demonstrated capability in thinking clearly, logically and critically in the paper is crucial for avoiding an outright rejection. By contrast, whether an adequate amount of detail is provided decides whether the paper it is of high quality or needs major revision. Although this variation was demonstrated only tentatively in the data, showing 'logic' and 'critical thinking' is the key to keep a paper in the reviewing cycle.

The requirement for critical thinking and analysis also applies to the evaluation of the literature review. If the author just 'compiled' studies without analysing them critically or they have not given appropriate credit to all the relevant literature, it can be viewed as being unacceptable. This specific can be deemed as an irremediable

issue within 'literature review' as a Group-A feature in general. By contrast, the issue of including unnecessarily too many studies or quotations is not a reason for outright rejection, but it should be avoided in a paper of higher quality. Compared to the other two problems, the problem of redundancy appears to be more remediable.

When it comes to writing, both the interview and review data have confirmed that a well-written paper has to be free of technical errors and jargon and use English and scholarly styles appropriately. While writing is clearly a Group-A feature and there is always room for improvement, the quality of writing has to reach a satisfactory standard when the paper is submitted. It can severely offend reviewers if too many grammar or spelling errors are presented. What indeed annoys them is the author's attitude behind such errors. Such authors are viewed as being unprofessional and unethical by exploiting reviewers inappropriately. The reviewers did show sympathy for non-native English speaking authors, yet expected such authors to seek assistance before submission.

Fit for journal appears to be a unique feature. Unlike the other Group-A features, it is judged dichotomously, which makes it appear to be more like a feature related to irremediable problems. However, the reviewers would always offer suggestions to amend the paper to make it fit for the submitting journal or suggest alternative, more suitable journals. In this regard, fit for journal is more remediable than not.

The comparison revealed variations between the specifics of the four Group A features. The variations uncover 'thresholds' in judgments, in a sense that the specifics vary in their potential to discourage reviewers' interest in, and patience for, a paper.

The specifics with the greatest negative influence on reviewers' judgment were: lack of critical and logical thinking in the argument and the literature, neglect of key literature, and having basic grammar and spelling errors through the paper. The other problems were less likely to attract outright rejection, but would not be expected to be present in a higher quality paper to the right side of the Line of Excellence on the Quality Continuum.

8.9.2 Group B: Features related to irremediable problems

Throughout the analysis of the interviews and reviews, Group B features 'novelty' and 'contribution' were separated into two categories, for the reason that the reviewers always commented on them separately. However, the analysis of the interview data revealed that novelty and contribution both had something to do with 'originality'. At the end of the analysis of the reviews, a word search for 'originality' was conducted. Only two reviews, both from education, actually used the word, writing:

> I found the concept of the 'conduct of concern' highly original and resonated strongly with my own experience of academia. [1249-3]
>
> This paper is a radical and highly original theoretical and methodological invention into research as usual in the field of education… makes innovative use of philosophers/theorists who have not been taken up much in education. [1255-1]

Both comments noted higher quality as being 'innovative' and 'making me think'. Other than these two, the reviewers did not cite the word 'originality' directly in the reviews, nor did they discuss much about it in the interviews. The literature strongly suggests that originality is a key feature in the evaluation of quality in grants (Guetzkow et al., 2004). In this study of manuscript review, the analysis shows that reviewers embedded their expectations of originality in the comments on contribution and novelty.

The feature of 'novelty' and its two specifics can be immersed into 'contribution'. The two features made a paper exciting and elicited a 'Wow' from the reader. They indicate what originality, or cutting-edge research, to use the reviewers' words, should look like. By identifying originality in certain papers, reviewers play an important role of defining the boundary of knowledge and pushing the cutting-edge forward.

Regarding the conception of 'originality' and 'cutting-edge', the education reviewers focused more on the choice of topic of the research and its relevance to practice, while the physics and chemistry reviewers attended more to research approach and findings. This indicates that, while in physics and chemistry the presence of new

findings or new application of experimental approaches can be detected relatively easily, in education whether a study is cutting-edge often has to be gauged by more subjective measures, for example, whether the research topic is 'timely' and 'relevant'.

The other Group B feature, research design and methods, did not received much unsolicited discussion in the interviews or reviews. Instead of commenting on methods directly, the reviewers often encouraged authors to provide more detail to allow for a better understanding of the methods. This implies that the credibility of the research depends not only on what is reported but how it is reported. One of the aspects the researcher now regrets is not exploring further the lack of comments on method in the follow-up email survey.

This section concludes the analysis and discussion about reviewers' considerations, approaches, and language used in the evaluation of research papers. Chapter 9 will bring together all the discussions reported so far and provide an integrated interpretation and set of conclusions for the thesis.

CHAPTER 9: CONCLUSIONS AND DISCUSSION

9.0 Introduction

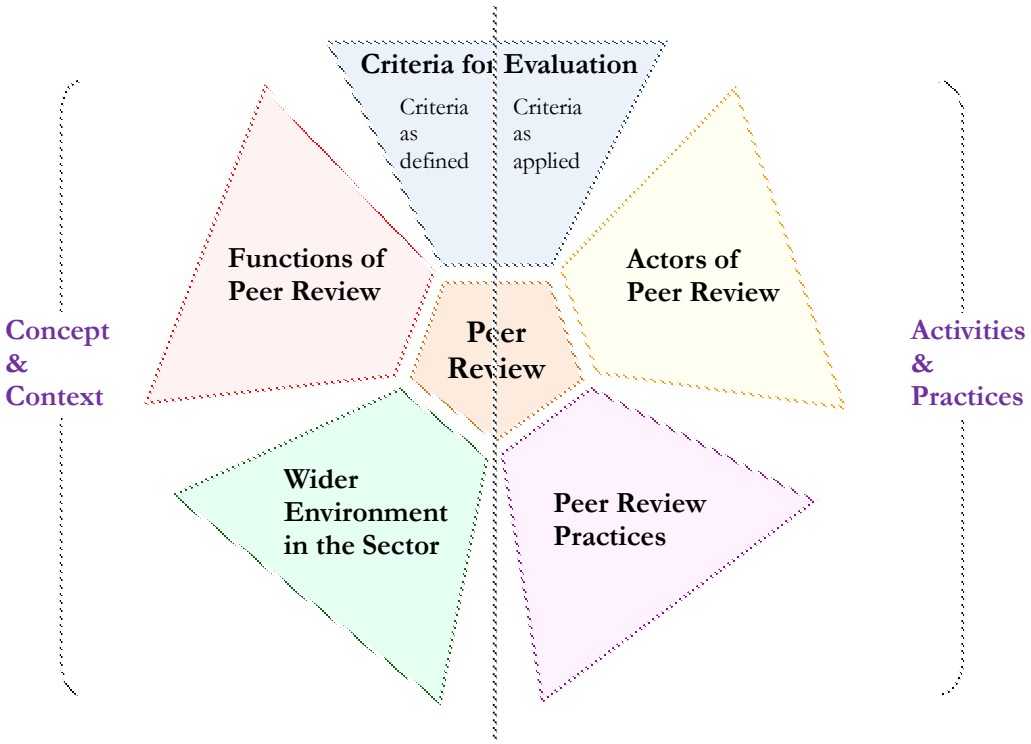
It is testimony to the strong interest in journal peer review that a number of studies have been undertaken in the area since the data collection and analysis for this study began. Among the most recent is an international cross-disciplinary survey (Mark Ware Consulting, 2008) that explores academic attitudes toward peer review and the experience of the process around issues such as anonymity, duration from submission to publication, and usefulness of reviewer comments. Initially surveys were sent to 41,140 academics and 3,040 replied (a response rate of 7.4%), and 95 percent were current authors. Of them, 85 percent recognised peer review as a valuable component of scientific communication, and 90 percent believed that it was effective in improving the quality of submissions. While many agreed that peer review could be improved, only 12 percent indicated they were dissatisfied with it, and 32 percent felt strongly that the current system was the best that could be achieved. This survey confirms the acceptance of peer review as a valued component of the publishing system as revealed in many earlier studies, but it also shares some of their shortcomings, namely the relatively narrow perspective taken by the researchers and a limited statistical analysis of the data. The international element of the survey design was novel, but little expanded.

A study of the literature undertaken for this thesis allowed the researcher to elaborate the functions of journal peer review, and it also highlighted the multiplicity of facets of peer review. In this context, the researcher has followed the aim of understanding 'what people mean and intend by what they say and do… within the historical, cultural, institutional, and immediate situational contexts that shape them' (Moss et al., 2009, p. 502).

From the literature, a working framework of five core dimensions that contribute to a comprehensive understanding of peer review was identified. The dimensions include the functions of peer review, its practices, actors, criteria for evaluation, and

environment. Figure 9.1 illustrates the framework. The diagram shows that the phenomenon can be differentiated into those dimensions or aspects of dimensions that capture the concept and context of peer review and those that capture its activities and practices. The Figure shows the complexity of the phenomenon and illustrates that an examination of any aspect of peer review would benefit from recognizing these dimensions, and a more holistic approach.

Figure 9.1 The Five Core Dimensions Contributing to the Understanding of Journal Peer Review



Coming to the study with an interest in research quality and its assessment and guided by this framework, the study investigated three questions about peer review using a mixed methods design of six phases that ultimately led to a fine-grained analysis of the distinctions made by reviewers when judging quality, so contributing a fuller explanation of judgments in different disciplines. The questions were:

i) How are the principles and practices of peer review defined within education and the physical sciences?

ii) How do reviewers perceive and address their role and activities in their respective disciplines?

iii) What judgments are made about quality and why, and do either differ by discipline?

This chapter begins with a discussion of the contribution of the study; it proceeds to highlight the most significant findings, then notes the limitations of the investigation and concludes with directions for further research.

9.1 Contribution of the Study

Three principles were adopted in the design of the study which addressed areas that were little studied in the literature: the first was to treat journal peer review as a holistic and complex phenomenon; the second was to use an approach that compared contrasting disciplines (education, physics and chemistry); and the third, to place a major emphasis on the knowledge of experienced reviewers in respect of evaluation and judgment of research quality. The latter specifically addresses the questions raised in the literature about a lack of clarity among scholars in articulating 'what high-quality or good research might be' (Forbes, 2003, p. 272).

Reviewers have often been the subject of research but rarely have their understanding of what is involved in judging quality been explored, especially by interview. In this study interviews with experienced reviewers bracketed by two surveys and followed by an analysis of their reviews were adopted to probe deeply into such understandings in the context of their perceptions and experience of reviewing.

A number of scholars have debated whether the principles of peer review translate equally across disciplines. Yates (2004) noted that the way in which research was perceived and assessed in education would be different from that in experimental sciences and as a consequence would be treated differently. Moss et al. (2009, p. 502) argued that, even within the social sciences, 'different perspectives on the aim of social science can entail substantially different perspectives on quality and rigour' and a 'unified' conception is hard to achieve.

Yet the question 'how different' the perspectives are in the context of journal peer review has not been substantially addressed, and direct comparisons of very different

disciplines are rare. To make the point that education is 'different' educational researchers often use the hard sciences as a point of contrast in epistemology and therefore 'judgment' about what constitutes quality (e.g., Moss et al., 2009).

Through its findings, the study challenged the common perception that there are major differences between disciplines in terms of expectations and judgments of research. There were more similarities than differences found across disciplines not only in practices, but in the specific considerations of quality and the language of review. A summary can be found in Appendix 16, across journal policies, survey and interview data, and analysis of reviews.

The study also explored published sources of information that have received very little attention, namely, journal policies and editorials about peer review. These documents tap into a range of areas in all the dimensions outlined above. Policies are published on journal websites; they contain details about the journal and publication in the journal including author requirements and peer review procedures. The examination of the policies of top ranking journals in education, physics and chemistry identified few disciplinary differences at this level as well.

The analysis of editorials examined the scholarly expectations and philosophy of editors about peer review spanning disciplines and journals. It revealed that editors were explicit in their expectations of reviewers' responsibilities, but less so about the qualities they sought in the reviewers. Editors attributed most problems in peer review to the 'submission explosion' and many admitted the difficulty in making major changes to the system. Also explored was the debate on open peer review, and the researcher developed a 'Spectrum of Secrecy-openness' (Figure 4.1, Chapter 4) to illustrate different editorial practices in this regard.

This study has provided a comprehensive account of the functions of peer review from the literature that shape practices, expectations and role. The three functions that date back furthest in history – controlling quality, providing expert advice and building credibility of published material, remain the primary functions for contemporary journals, while improving quality, stratifying submissions, defending scientific autonomy, and nurturing collegiality have emerged more gradually.

9.2 Main Findings of the Study

In this section, the main findings are presented under a number of headings. The first five relate to reviewers and their practices, namely: reviewers' expectations of peer review; their understandings of the task of reviewing; criteria emphasised in making judgment on quality; thresholds at which reviewers differentiate manuscripts by quality; and the language of review. The following four sections comprise its functions; the regulatory landscape; underlying causes of the problems besetting peer review; and broader environmental influences on peer review.

9.2.1 Reviewer expectations of peer review

Reviewers perceive two features as crucial for the ongoing commitment of reviewers and high quality reviews. They are an appropriate level of feedback from the editor regarding their reviews and an appropriate level of recognition of their contribution by their employing institution.

There are at least five types of editorial feedback, and the respondents have shown very different attitudes towards the necessity of providing feedback of different types. Most of the reviewers expect that they will be notified of the final decision and feel offended if the editor fails to do so. They regard the acknowledgement of receiving the review and a 'thank-you' message as an important routine that all journals should maintain. In contrast, they are not averse to receiving feedback on the quality of their reviews and perceive dealing with such feedback to be too time consuming.

The formal recognition of reviewers' contribution is keenly desired by reviewers. The study has revealed that reviewers generally believe that the activity of reviewing is respected by their colleagues, and the literature shows that reviewers receive a range of benefits (for example, Braun & Bujdosó, 1983). What they want is a higher level of recognition by their employing institution. Since reviewers perceive reviewing as a natural part of their academic life, they expect it to be treated, and recorded, as a formal academic activity, in the same ways as teaching and supervising.

The issue of overloading qualified reviewers, and the pressure on all academics, has exacerbated the need for the appropriate recognition of reviewers. Many of the

respondents believed recognition will win back many qualified reviewers who have declined to review because they believed their input would not bring in an equivalent return. As a result, many problems connected with a lack of adequate number of reviewers in the current system, such as difficulties in matching a work with appropriate reviewers and unhelpful comments, may be diminished.

9.2.2 Reviewer understanding of the task of reviewing

Reviewers embark on the task typically with a reasonable amount of research experience and publications, but no formal training for reviewing. Learning to review is primarily a self-guided process, with the engagement of limited mentoring, assistance and explicit guidelines. New reviewers adapt skills from related academic experiences such as marking student assignments, reviews of their own PhD theses or publications; and they seek advice from supervisors, colleagues, and journals. This process of finding one's own path in academic growth – the self-guided path and the associated cycle of learning, is common in academic culture.

New reviewers can feel a lack of confidence about reviewing for an extended period. The feeling is directly related to their perception of how established they are as a researcher. This awareness involves their familiarity with the field, and experience in academic writing and evaluating academic writing. They gradually establish a personal reviewing style and approach through trial and error and continuing confirmation of their success as a reviewer. Reviewers do not appear to reject this process of trial and error, as in general they are not in favour of the introduction of formal training, believing it to be unnecessary and impractical. What is emphasised by the majority of the respondents, however, is that journals can play a crucial role in producing good reviewing by specifying unambiguously their expectations of what a 'good review' is.

Rigorous peer review costs significant time and effort, not to mention most reviewers volunteer this. Literature shows reviewers are driven by a concern for the quality of their field and the journals they read and publish in (Walker et al., 2008). Reviewers gain reciprocal benefit by reviewing for high quality journals (Engers & Gans, 1998), and favour this above monetary reward (Tite & Schroter, 2007). The

respondents to this study view journal peer review primarily as a professional obligation and an expectation of them as academics, followed by the opportunity it offers for professional development and collegial reciprocity. This mirrors the notion of disciplinary 'stewardship' (Walker et al., 2008) and the argument that 'reviewing is at its most powerful when it works to transform education and research, and that it can only do that by transforming the actor' (Graue, 2006, p. 36).

Only a few respondents cited improving paper quality as a primary motivation from their perspective as a reviewer; however, most of them anticipated an improvement of quality as an outcome of peer review; they also rated the necessity of this function very highly in the initial survey.

In Australia, there is another important application of peer review – evaluation of grant applications for the Australian Research Council (ARC). This study showed that respondents were significantly more satisfied with journal peer review than grant peer review in all aspects studied, especially when they considered their experience as authors or assessors. Comparison between the two practices highlights the instructive and constructive nature of journal peer review compared to grant review – when acting as journal reviewers the respondents felt more comfortable helping authors improve their submissions than when they were grant assessors.

9.2.3 Quality features in the consideration of manuscripts

Macnab and Thomas (2007, p. 340) indicate that there are problems of 'using criteria as general markers for quality'. This study provides insights into the criteria used for evaluation. The seven groups of criteria identified in this study are: writing, argument, literature review, fit for journal, research design and methods, contribution, and novelty. The broad grouping of the features found in this study is not that different from what is documented by other researchers interested in this area (for example, Bakanic et al., 1989; Campion, 1993; Fiske & Fogg, 1990; Gosden, 2003) despite the fact that the researcher attempted to elicit new angles on judgment.

It is not possible to escape that reviewers think in terms of these functional categories. However, what this study contributes that is different is an account of

reviewers' subtle considerations of these features, as well as providing better focus on disciplinary variation.

As also revealed in previous studies in other disciplines (Bakanic et al., 1989; Fiske & Fogg, 1990), this study has confirmed that 'argument' is the most heavily weighed feature in the evaluation for both education and physics and chemistry reviewers. The strength of argument lies in its clarity and consistency. 'Clarity' involves the requirements of presenting sufficient detail to explain key points, organizing the discussion in a logical order, critical thinking and analysis, and avoiding overstatement. 'Consistency' is about how well the conclusion is supported by the evidence and how well this link is demonstrated in the paper.

'Contribution' is the second most emphasized feature. It embraces: topic choice, relevance to practice, advancing knowledge by producing important findings and advancing knowledge by adopting a certain approach. There is a marked disciplinary variation in reviewers' emphasis on these specifics. The education reviewers privileged 'topic choice' and 'relevance to practice', while the physics and chemistry reviewers prioritized contribution in the form of 'adopting a certain research approach' and 'providing important findings'.

With regard to a good 'literature review', both the quantity and quality of articles being quoted in a paper have to be appropriate. A lack of selectivity in the inclusion of studies or quotations or failing to give appropriate credit to all the relevant literature are perceived as demonstrations of poor quality. It is also perceived to be unacceptable if the author just 'piles up' studies without analysing them critically.

Another expectation is that the literature review should place the work in context. The education respondents emphasised 'making sense' – using the literature to situate and support the author's opinions and assertions presented in the work, while for the physics and chemistry respondents it was more about providing sufficient background for readers to better understand the significance of the work. In addition, the education reviewers discussed the review of the literature at greater length in the reviews they presented than those in physics and chemistry.

Another feature that draws more attention from education reviewers is 'writing'. A well-written paper has to be clear, concise, fluent, using English and scholarly style appropriately, free of technical errors or jargon and comprehensible to non-specialists. It is crucial that the quality of writing reaches a satisfactory standard when the paper is submitted, as reviewers tend to associate frequent errors in writing with the lack of academic rigour of the author.

'Research design and methods' did not receive much emphasis by the responding reviewers, except for a small number of chemistry reviewers pinpointing cases of 'wrong science'. Instead of commenting directly on the quality of method, the reviewers more often required authors to provide more detail to allow for a better understanding of the methods. Peer review works on a mutual trust between authors and reviewers (Smith, 2006a). This principle seems to figure prominently in the reporting and evaluation of methods – the credibility of the research depends not only on what is reported, but also how it is reported.

'Fit for journal' refers to the suitability of a paper to the focus, readership, and style of a journal. The requirement is typically included in the journal policy statements of 'relevance and impact' and 'following journal guidelines'. The reviewers indicated in the interviews that they would reject a paper if it did not fit the journal. The reviews prepared by some of the reviewers, however, revealed a more generous intention in practice. They either suggested alternative journals for the paper to try or provided suggestions for authors to amend the paper to make it 'fit' for the submitting journal.

The least mentioned feature is novelty. In the analyses of interviews and reviews, 'novelty' and 'contribution' were split into two categories, since the reviewers always commented on them separately. Yet the analyses show that they both have something to do with 'originality'. In journal policies, the words 'new' and 'novelty' recur in the statements for contribution. The simultaneous presence of the two features makes a paper exciting and elicits a 'Wow!' from the reader. By identifying originality in the submissions, reviewers make an important contribution to their field by defining the boundary of knowledge and pushing the cutting-edge forward.

There is also a disciplinary variation in the perception of originality. While education reviewers focused more on a study's topic choice and relevance to practice, physics and chemistry reviewers attended more to research methods and findings. This indicates that, while in physics and chemistry the presence of new findings or a novel application of experimental approaches can be detected relatively easily, in education whether a study is 'cutting-edge' has to be gauged by more subjective factors, for example, whether the research topic is 'timely' and 'relevant'.

9.2.4 Distinction of manuscripts by quality

In the literature there is 'a complete lack of specificity' about the nature of good quality research (Forbes, 2003, p. 272) as well as the 'exact level of cumulative errors' at which reviewers reach the recommendation of rejection (Kassirer & Campion, 1994, p. 96). This study has revealed that while reviewers generally follow the same broad groups of criteria in reviewing, they weigh the relative gravity of various flaws differently and draw on different combinations of quality features when making recommendations. The study has also identified the 'lines' along which different reviewing decisions are drawn.

A Quality Continuum has been developed to illustrate the distinctions of quality by the line of 'Remedy', 'Excellence', and 'Impact' (see Figure 7.5 in Chapter 7). The development of the Quality Continuum was informed by the five ways of thinking about quality identified by Harvey and Green (1993, p.10), who contended that the concept of quality 'is not a different perspective on the same things but different perspectives on different things with the same label'. This study confirmed that, under the same label (namely, quality features), there are more subtle considerations.

The 'Line of Remedy' marks the distinction between outright rejection and marginal acceptance; the 'Line of Excellence' marks that between marginally acceptable and good quality; and the 'Line of Impact' indicates a higher level again: papers passing beyond this line are bound to show impact sooner or later; it marks a major transition from good to extraordinarily high quality. Towards the far right end of the Quality Continuum, there are features that denote the 'best' research paper.

A paper that is located at the left side of the 'Line of Remedy' will be rejected outright. Such a paper often fails to demonstrate the acceptable standard required by the field served by the journal. Flaws include that authors fail to show capability in thinking clearly, logically and critically in the argument and in the literature review; they ignore key studies on the topic; and there are too many grammatical or spelling errors. Reviewers are highly critical towards the latter – they view a sloppy author as being unprofessional and unethical by inappropriately exploiting the reviewers.

Such a paper also often uses inappropriate or 'low level' methods; fails to support the conclusion by the evidence reported; delivers nothing new; and fails to make sufficient contribution to the field. Physics and chemistry respondents in particular often cited 'lack of novelty' as the reason for outright rejection.

When determining qualified acceptance as opposed to outright rejection, those interviewed highlighted clear and logical research method, contribution and novelty (by using the word 'must' repeatedly) as features they expected to find in the paper. The paper free of those flaws may be considered for publication, but still receive a recommendation of 'major revision' and need substantial additional work to fix flaws typically in the coherence and completeness of the argument, appropriate inclusion of literature, and in the quality of the presentation of the paper.

A paper will be located towards the 'Line of Excellence' if it has clearly met all the above expectations. Reviewers pinpoint 'high quality' in two ways. The first, 'all and better', recognizes the presence of all the quality features, but at an elevated level and has been cited much more often by education reviewers in the interview responses. The second focuses on a single outstanding feature in a paper that has met all the requirements. Reviewers agree that 'novelty' is that feature (even though they rarely present detail on this) and indicate such a paper makes them think in a way they have never thought of before. Many also emphasise 'well-written argument' which makes the novelty of the work easy to understand. This latter feature is much more often cited by physics and chemistry reviewers.

While many of the reviewers admitted that they would accept a paper as high quality even if it was not that ground-breaking, they did expect to read something that

convincingly changes the way people think about certain issues and even redirect the field. Such a paper will be positioned at the 'Line of Impact' if it is not only 'excellent' in itself, but also of great social or scientific benefit to, and showing strong sense of usefulness or direction to, other researchers in that field.

Placement at the far right end of the Quality Continuum is rare. Most of the reviewers interviewed indicated that they had never reviewed or read a paper that could be called the 'best', and their responses were often accompanied with a tone of regret. Reviewers struggled to describe the quality of a 'best' paper. Their responses highlight an expectation for all features of top quality and 'superiority'; such a paper has to come from a superior understanding of the phenomenon, which is typically gained through years of hard work and immersion in a field.

9.2.5 The language of reviewing

By 'language of reviewing', the researcher refers to the language of judgment and how such judgments are communicated. Previous studies of reviews consistently show that the predominating tone is negative (for example, Bakanic et al., 1989; Fagan, 1990; Fogg & Fiske, 1993; Runeson & Loosemore, 1999; Tight, 2003).

This study confirms the dominance of negative comments in general, but it also reveals that reviewers comment 'more negatively' on some quality features than on others. It identifies two distinct groups of features (referred to as Group A and Group B for convenience). Group A are related to remediable flaws and Group B are related to irremediable flaws. The grouping emerged from the characteristics of the comments used in the reviews.

Group A includes: argument, writing, literature review, and fit for journal. Reviewers' comments on the first three features share at least two characteristics: they come to a judgment that reflects the 'extent' to which a paper satisfies the quality they expect, and always signify that there is room for improvement. In the case of 'fit for journal', a paper can be judged to be unsuitable for one journal and rejected for it but still be acceptable by another, thus making it remediable in a broader sense.

Group B includes: research design and methods, contribution, and novelty. Flaws detected in this group are not (easily) salvageable in their present form, and the judgment on them is typically constructed in a dichotomous manner, yes or no.

Three clear patterns in reviewer comments are evident. First, negative comments on Group A are much longer in length than those on Group B: reviewers usually provide extensive explanations of the problem and suggestions for improvement for Group A, while such details are largely absent from comments on Group B.

Also for Group A, positive comments are much shorter in length than negative ones. When high quality is demonstrated, reviewers often only acknowledge it briefly. If a problem is detected, they will then explain in detail why it is a problem and how it can be addressed, often by asking questions and providing suggestions. For Group B, negative comments outweigh positive ones (as with Group A), but more detail is provided on positive features for example by explaining why the paper has made a significant contribution.

The quality features can be further stratified into 'specifics' (see Appendix 15 for detailed grouping). For Group A, such specifics are well identified in the interviews, and the review data substantiated them with more detail, such as the four meanings embedded in the 'clarity' of argument. This indicates that the reviewers established their conceptualizations of the features in a generic, stable and solid manner. By contrast, when discussing Group B features in the interviews, the reviewers provided only broad and brief responses. Reviewer comments varied widely in focus; the way the comments were constructed reflects an immediate recognition of the features but a difficulty in pinpointing and articulating them.

The analysis has also revealed some disciplinary variations (also see Table 9.1), which were less distinctive than those distinguished between Groups A and B. Education reviewers commented more often on 'literature review' and 'writing' than physics and chemistry reviewers, and the reviews from education were much longer in general. In addition, when commenting on 'contribution', their emphases were different – education reviewers focused more on 'topic choice' and 'relevance to practice' and physics and chemistry reviewers more on 'findings' and 'research

approaches'. Other than these, there appear to be more similarities than differences between the two disciplines in the interpretation and evaluation of 'quality'.

Reviewers' perceptions and practices do not exist in vacuum, but are affected by a number of factors, including belief in the contribution of journal peer review, regulations given by journals, problems in peer review, reviewers' own expectations of the task, as well as the wider environment in the sector. The following sections discuss findings in regard to these factors.

9.2.6 The functions and contributions of journal peer review

The basic concept of journal peer review, that external experts provide advice to assist editors in the selection of material for publication, was established several decades ago by the earliest scholarly journals Philosophical Transactions and Journal des Scavans (Kronich, 1990). These journals introduced the system for the purpose of controlling quality, providing expert advice and establishing credibility (Burnham, 1990). More recently recognised functions include improving the quality of papers, stratifying papers by quality, nurturing collegiality, and defending scientific autonomy (Beyer et al., 1995; Cain, 1999; Knoll, 1990; Liesegang et al., 2005; Osborne & Brady, 2002).

Editors express strong faith in peer review's contribution to scholarly publication in editorials and journal policies, not only as a tool of quality control, but also as a means to bring the research community together. Editors expect reviewers to be responsible to at least four groups. For readers, reviewers select the 'best' for publication and ensure what is published is worthy reading, which reflects the functions of 'quality control' and 'establishing credibility'. For authors, reviewers have a collegial obligation to act ethically, return reviews promptly, and to provide constructive comments in pursuit of 'improving manuscript quality'. For editors, reviewers share editors' workload and broaden their knowledge on which editorial decisions are based, which reflects the function of 'providing expert advice.' And for the research community, peer review contributes to 'nurturing the collegiality' of a discipline across institutional and geographical boundaries and 'defending scientific autonomy'.

A slightly different yet compatible conception of the functions of journal peer review is revealed in reviewers' responses in the interviews (Figure 6.1 in Chapter 6). At the operational level, peer review is an iterative process which involves a back and forth dialogue between author, reviewer, and editor. At the functional level, peer review is a tool of gatekeeping, maintaining high standards, and establishing authority and credibility of scholarly publications. At a more personal level, peer review assists in building capacity of researchers, as well as improving the quality of individuals' work. These functions can be further grouped into immediate contributions to journal publication and prospective contributions to scholarly communication and the wider research community.

The study confirms that reviewers and editors generally value peer review more for its contribution to scholarly communication over its contribution to individuals. 'Maintaining high standards in publication' is the most expected function, followed by 'gatekeeping' and 'establishing credibility and authority' of published work.

9.2.7 Expectations and guidelines published by journals

Scholarly journals all follow certain procedures for reviewing and publishing. A considerable proportion of journal policies analysed in this study focus on explaining this. The reviewing processes adopted by journals are very similar. It typically starts with acknowledging the receipt of a paper, and continues by passing the paper through pre-screening and external reviewing, notifying authors of the initial decision made on the basis of the returned reviews, and assessing the revised paper to make a final decision. The process concludes by logging accepted papers in the waiting list for publication or dealing with author appeals. The overall process is handled by editors and their assistants.

The greatest variation in the process is in the selection of reviewers. This study confirms that physics and chemistry journals assign fewer reviewers to each paper than in education, and they have a common policy of inviting authors to suggest and exclude potential reviewers. Editors from different journals or fields can have very different expectations in the selection of reviewers. The grounds on which reviewers are chosen do not receive much attention in the policies or the editorials. A

sociology editor once noted that the quality an editor looks for in a reviewer was competence, which is 'an obvious answer but one that says too little about a concept with many levels of meaning' (McCartney, 1973b, p. 290).

In editors' considerations, the 'multiple levels of meaning' may include expertise, availability, the likelihood of providing unbiased and timely reviews, free of conflicts of interest, and a balance in expertise, nationality, and institution. They are complex and can make reviewer selection appear to be ambiguous. The initial survey identified many cases where the reviewers were sent papers that were outside their area of expertise.

The practice of anonymous/blind review is less a focus of journal policies. When it is mentioned, the meaning of the terms is not always well defined, with 'blind' and 'anonymous' often being used interchangeably. A general trend identified in this study is that physics and chemistry journals are more likely to use single-blind review, and some even discourage reviewers to release their identity.

What is the value or necessity of anonymity? This study confirmed reviewers' commitment to anonymity that was revealed in the literature (Bingham et al., 1998; Fabiato, 1994; Fontaine, 1995; Posey, 2005). Most of the reviewers anticipated more harm than merit in an open system. They perceived electronic journals that more often used open peer review as of lower quality than traditional print journals. Even those who were somehow supportive of an open system maintained that the choice of being anonymous or not should always remain with the reviewers.

Extensive discussions on this topic are also found in the editorials. In the presence of a few promising trials of signed review policies (conducted mainly by medical journals), most editors have conveyed a conservative attitude – in the editorials they perpetually weigh its strengths against its weaknesses; they encourage readers to comment on the options, but at the same time choose to continue with anonymous review to wait for 'stronger evidence for an open approach'.

9.2.8 Problems in journal peer review and potential solutions

Problems in peer review, as those related to its fairness, effectiveness, and efficiency,

are well documented in the literature. Editors of different journals, fields and time periods have identified similar problems, which are mainly related to three factors: paradigmatic consensus in a field, scientific ethics and integrity, and submission explosion. The lack of paradigmatic consensus in some fields can place editors in a difficult situation when it comes to matching reviewers to authors, reconciling disagreements between reviews, or identifying reviewers for an impartial evaluation of nontraditional work. The inflation of submissions has intensified the tension between editors and authors and increased concerns about sub-standard reviewing, and less than optimal decision-making by editors.

Finally, on ethics and integrity, a number of mechanisms have been tried that seem to present a system of totally opposing views, from double blind peer review to open reviewing; and the use of more reviewers for each paper to achieve more balanced views to the use of fewer reviewers to maximize the use of strong reviewers and reduce publication lag and reviewer workload. None of the actions has 'overhauled' the process and many issues discussed three decades ago are still worrying the editors of today.

There has been no system-wise change to peer review, and problems with peer review were volunteered by the participating reviewers in this study. They indicated a high level of satisfaction with the current system, irrespective of discipline, gender or academic status. Nearly half of them indicated that no change was needed. The rest pointed out areas where changes might be necessary such as those related to the efficiency of the process, but most of them did not offer any feasible strategy to realize the change. No one suggested replacing peer review with any alternative. This position held even when author attempted to probe further on the issue during the interview and follow-up email survey.

So, is peer review 'immune' to change? Otherwise why are changes to peer review so difficult to make? Glenn (1979, p. 785) once argued that for peer review there is 'no effective pressure for change'; 'any suggestion for innovation is always countered by the charge that 'it won't work' – usually in the absence of any empirical evidence'; those who are 'best able to promote change' are 'not very motivated to do so' and

the 'young, untenured… are unlikely to… demand reform.' More recent literature has not revealed any change to the situation. The findings in this study confirm what Glenn had noted so many years before, that the experienced, established scholars, those who are best able to promote change, are not clearly or highly motivated to promote change.

The participants specifically noted concerns regarding efficiency, editorial leadership, adequate explanation of criteria for evaluation, feedback to reviewers, appropriate recognition of reviewers' contribution, as well as problems associated with single blind review. What was conveyed in parallel with the criticisms, however, was a strong sense of faith in peer review and this pervades editorials as well. When editors and reviewers were critical about peer review, they often emphasised the indispensable status of it right after a criticism was made or noted that the problem was related to individual idiosyncrasies while the system was fine. Many of the participants used the free comment options in the survey or the interview as a chance to 'soften' the critical remarks they had made previously with an expression of faith. The participants demonstrated a very high level of tolerance of its deficiencies and this in turn may be related in part to the awareness that for most of the problems there is no obvious or feasible solution, and in part may arise from an appreciation for the voluntary, unpaid nature of the task.

The co-existence of criticism and faith reflects perceptions identified in Chapter 1 that much like democracy, peer review functions as the 'worst possible system except for all others'. Decision-making is exercised through the separation of powers between editors and reviewers, and decisions are made on the basis of a collaboration of expert knowledge through a process of negotiation of diverse opinions and considerations. Where criticisms emerge is that editorial decisions are not always based on a consensus among the majority, this is increasingly being countered, however, by arguments that divergence among reviewers is an asset not a problem.

With the nature of the problems, the question is not whether or not to change but 'how' to change. There will be no revolution from within. Continuous quality

improvement (CQI) can be a feasible approach to enhance the process. CQI, better known in the industry as Total Quality Management, is a management philosophy that has made stronger inroads into the higher education sector since the early 1990s (Smith, 1997; Wild, 1995). CQI is a democratic approach and is compatible with peer review's pursuit of academic collaboration and autonomy, rather than imposing solutions on the people who are involved in the process it calls upon their support and expertise to achieve the improvement. In pursuit of a successful continuous improvement, it is important that journals and editors create a climate where reviewers, authors and readers are continuously encouraged to criticise the status quo and suggest improvements, where the authors decide where quality is improved, and where a mechanism is in place to gauge and compare the outcomes continuously.

9.2.9 Influences of the environment

As a core activity in academia, peer review is affected by and reflects changes occurring in the wider academic environment, most notably the 'pressure to publish'. In Australia, the pressure has been exacerbated by the evolution of national quality assessment schemes, from the Institutional Grants Scheme (IGS) to Research Quality Framework (RQF) and to Excellence in Research for Australia (ERA). As a result increasing weight has been placed on publications and the ranking of journals that will define the nature of academic activity in Australia for many years to come, influencing the conditions, rationale, and priorities behind academic work, and by association the conditions of peer review. Participants in this study noted that in the 2000s, the pressure to publish had increased dramatically both in Australia and globally because publication generated income for their workplace.

There are two consequences of this in the way in which academics go about publishing. One is that they are more likely to submit work of substandard quality, in an attempt to just get a lot of work out for review. The other is they expect that their submissions must get accepted, and any failure to publish can cause significant stress.

Reviewers are also victims of the pressure to publish. The pressure to publish has led to a marked growth in the number of papers submitted without marked growth in the overall quality of the submissions (Bedeian, 2003). Exhausted by the growth

in the demand for reviewing and the number of substandard papers, reviewers can become less inclined to offer their availability. And they are also authors and need to allocate considerable time to research and write for publication. Many of the responses reflected reviewers' struggle between the awareness of collegial obligation and their individual availability.

Some other trends include electronic submission and reviewing as stimulated and enabled by IT development, and a general better organisation of peer review system in journals due to the pressure of journal ranking practices. Their long-term influences are yet to be seen.

The only main area in peer review that has not been apparently affected by the external environment are the criteria for evaluation, or, the considerations and interpretations of quality. None of the respondents in this study noted any change in the criteria for evaluation over time, and the quality features identified by this study and by previous ones are pointing in the same broad directions.

9.2.10 The framework for the understanding of journal peer review

The findings discussed above are pertinent to the functions, practices, actors, criteria for evaluation and environment of peer review. These five core dimensions constitute the framework for a comprehensive understanding of peer review. The framework is presented as Figure 9.1 at the outset. The findings confirm that the five dimensions are interrelated to one another closely.

For example, the intensifying pressure for publication (regarding the dimension of 'environment') has caused a shortage in the supply of qualified or willing reviewers ('actors'); the difficulty in finding sufficient reviewers would lead editors to relying more on in-house review or on authors suggesting reviewers ('practices'); and the emphasis in the evaluation of submissions ('criteria for evaluation' and 'functions') may change too, for example, from encouraging authors to improve their work to 'finding a reason to reject the paper', where the reason can be as vague as 'there is nothing new'. This relationship is also acknowledged by a recent article on peer review (Benda & Engels, 2010).

The framework indicates a promising way to conceptualise and address conditions that may threaten the soundness of journal peer review, as well as to improve the system. This is probably best achieved through continuous quality improvement.

The interrelationship among the dimensions also explains why changing peer review is so difficult. Often a proposed solution only attends to the problems in one dimension or one aspect of a dimension, while the implementation of it can cause new problems in another dimension. A good example of this is the introduction of an 'open' system. This initiative may make the process of reviewing quicker and appear to be fairer, but it may also lead to more reviewers declining to review or an unhealthy minimizing of the gate-keeping function of peer review. The consequences become particularly unaffordable in the current environment of academic pressure.

9.3 Weaknesses of the Study

During the course of the study, the researcher became aware of the weaknesses in this project and the endless potential for future research. They are discussed in the following two sections.

The researcher chose experienced senior academics as research subjects and they are by definition a busy group of professionals. This choice narrowed down the number of people who were able to participate. As a result, 84 responded to the initial survey (9% of the population of 930 'senior academics' or a response rate of 36.2% of the 232 people who were sent an invitation) and 44 of them participated in the interview. The design of the study attempted to maxmise the information that could be obtained from these experienced academics.

As far as the research instruments are concerned, a number of items could be added to better support some of the core findings. For example, the review of theoretical articles revealed a comprehensive list of 'functions of peer review'; this list could be added to the initial survey for the respondents to rate/rank their importance.

Two interview questions that generated interesting results, 'the stage of feeling confidence in reviewing' and 'the necessity of formal training for new reviewers'

were added to the protocol after a few interviews were conducted. If another study were to use the protocol, these questions and possibly more follow-up questions, should be included. Otherwise they should be better explored in the follow-up brief survey; among these are questions seeking to elicit responses that draw out more examples and illustrations because reviewers find it very hard to articulate their decision-making (for example, judgments of 'research method'). However in this study when the follow-up survey was sent with the interview transcripts to the respondents for checking, a tight timeframe had to be followed which left no time to determine the very specific areas that would require further elaboration.

In addition, the study has dealt with some material that changed quickly during the course of the project. For example, the ERA ranked journal list was just released in 2008 when it was used for the study, and was replaced by a revised list in February 2010. It was too late for this study to incorporate the new information. However, the researcher believes that given the scope and size of the list of journals contained in this study, little change would arise in the findings, although this should be the subject of examination in a follow-up study.

9.4 Directions for Future Research on Peer Review

This thesis contributes to the emerging literature on how judgments occur in the context of research assessment and offers a framework to evaluate the effectiveness of journal peer review, to detect imbalances in the system, and to inform reviewer development.

The multi-faceted nature of peer review across disciplines offers considerable potential for new research, especially in regard to how decisions are arrived at. The language, culture and intention behind assessment and evaluation of research in the modern academy are just beginning to be explored in detail (Forbes, 2003; Kiley & Wisker, 2009; Lamont, 2009; Lovat et al., 2008). Nonetheless, some topics in the practical areas of peer review have been quite exhaustively examined, for example reviewer agreement and bias on which many studies were published in the 1980s and 1990s. Others have just begun to attract attention such as the implications and nature of 'open' peer review. The social and psychological aspects of journal peer review as

well as other types of reviewing, for example, group-based peer-review (Benda & Engels, 2010) are also very much understudied.

Those who investigate peer review can be grouped broadly into 'journal insider' and 'outsider' categories. Most previous studies have been conducted by 'insiders' such as current and former editorial staff and researchers who are connected to the target journal in one way or another. More investigations to be undertaken by outsider researchers should be encouraged.

This study has proved that creativity in design benefits the study of peer review. The mixed methods used in this study is unusual in the peer review literature (possibly because so many studies are single discipline and even method-focused). In addition, the opportunity taken to revisit the reviewers in email mode proved productive and extended to deeper reflections by individuals on issues raised by the group as well as in relation to their own interview and survey response text. The growth in theorizing and practice of mixed methods is very pertinent to more complex questions being asked of peer review.

This study has compared reviewers' responses in interviews about what reviewers do in relation to identifying quality, and the comments they made in the reviews that they prepared. This approach has worked effectively in generating rich, validated findings. More studies of this kind, especially with a larger sample, are recommended.

In addition, this study has identified the finer meanings of the core criteria for evaluation. More quantitative, experimental studies can be built on this finding and be conducted with more reviews and in other disciplines.

Lastly, peer review is employed by journals of all disciplines. This study has chosen education, physics and chemistry as its focus and detected a small number of disciplinary differences regarding the evaluation of research manuscripts. Similar comparative studies can be conducted to advance knowledge in this regard and in turn can be used to inform the broad debate on quality, and to explore further the norms, conventions, and differences in this important area of higher education and knowledge production.

REFERENCES

Abbasi, K. (2006). JRSM introduces open peer review. *Journal of the Royal Society of Medicine, 99*(8), 379-380.

Abby, M., Massey, M. D., Galandiuk, S., & Polk, H. C. (1994). Peer review is an effective screening process to evaluate medical manuscripts. *Journal of American Medical Association, 272*(2), 105-107.

Abdou, A. (2004). Editorial. *Saudi-Arabic Medical Journal, 25*(Supplement), S43-44.

Abramowitz, S. I., Gomes, B., & Abramowitz, C. V. (1975). Publish or politics: Referee bias in manuscript review. *Journal of Applied Social Psychology, 5*(3), 187-200.

Abt, H. A. (1987). Are papers by well-known astronomers accepted for publication more readily than other papers? *Publications of the Astronomical Society of the Pacific, 99*, 439-441.

Adams, J. E. (1977). The challenge and response of economics journal policies: A comparative survey. *Collegiate News and Views, 30*(3), 25-27.

Alpert, J. S. (2007). Peer review: the best of the blemished? *The American Journal of Medicine, 120*(4), 287-288.

Alspach, G. (1994). What journal editors would like from reviewers? *Critical Care Nurse, 14*(6), 13-16.

Alstete, W. J. (1995). Benchmarking in higher education: Adapting best practices to improve quality. *ASHE-ERIC, Higher Education Report No. 5*. Washington, DC: George Washington University.

Altman, D. G., & Schriger, D. L. (2005). The statistical and methodology content of journals' instructions for authors. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Álvarez-Dardet, C., Gascón, E., Mur, P., & Nolasco, A. (1985). A 10 years trends in the Journal's publication. *The New England Journal of Medicine, 312*, 1521-1522.

Alvermann, D. E., & Reinking, D. (2004). Editorial guidelines concerning conflict of interest for Reading Research Quarterly (RRQ). *Reading Research Quarterly, 39*(1), 10-12.

Armstrong, J. S. (1996). The ombudsman: Management folklore and management science – on portfolio planning, escalation bias, and such. *Interfaces, 26*(4), 25-55.

Armstrong, M. (1999). Historical and contextual perspectives on benchmarking in higher education. In H. Smith, M. Armstrong, & S. Brown (Eds.), *Benchmarking and Threshold Standards in Higher Education* (pp. 7-34). London: Kogan Page.

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 Counselling Psychology, 29*(2), 189-194.

Atkinson, M. (1994). Regulation of science by 'peer review'. *Studies in History and Philosophy of Science, 25*(2), 147-158.

Australian Research Council. (2008). *New ERA for research quality: Announcement of Excellence in Research for Australia initiative*. Canberra: ARC. Available at http://www.arc.gov.au/media/releases/media_26feb08.htm

Autorino, R., Quarto, G., Di Lorenzo, G., De Sio, M., Damiano, R. (2007). Are abstracts presented at the EAU meeting followed by publication in peer-reviewed journals? A critical analysis. *European Urology, 51*(3), 833-840.

Babor, T. F., Edwards, G., Stockwell, T., & West, R. (1996). Inside Addiction: Peering into the editorial review process. *Addiction, 91*(12), 1757-1763.

Bachand, R. G., & Sawallis, P. P. (2003). Accuracy in the identification of scholarly and peer-reviewed journals and the peer-review process across disciplines. *The Serials Librarian, 45*(2), 125-158.

Baggs, J. G., & Minicucci, D. S. (2000). Shining a light on the black box of the review process at Research in Nursing & Health. *Research in Nursing & Health, 23*, 1-2.

Bailar, J. C. III, & Patterson, K. (1985). Journal peer review: The need for a research agenda. *New England Journal of Medicine, 312*(10), 654-657.

Baines, C. J. (1987). Reviewing the reviewers. *Canadian Medical Association Journal, 137*, 15.

Bakanic, V., McPhail, C., & Simon, R. J. (1987). The manuscript review and decision-making process. *American Sociological Review, 52*(5), 631-642.

Bakanic, V., McPhail, C., & Simon, R. J. (1989). Mixed messages: Referees' comments on the manuscripts they review. *The Sociological Quarterly, 30*(4), 639-654.

Ball, C. J. E. (1985). *Fitness for purpose*. Guildford: Society for Research into Higher Education.

Banks, J., & Pracht, C. (2005). Movers and shakers in the library publishing world highlight their roles: Interviews with print and electronic journal editors - A comparison. *Electronic Journal of Academic and Special Librarianship, 6*(3). Retrieved on 10 November 2007 from http://southernlibrarianship.icaap.org/content/v06n03/banks_j01.htm.

Barnett, A. H., Ault, R. W., & Kaserman, D. L. (1988). The rising incidence of co-authorship in economics: Further evidence. *The Review of Economics and Statistics, 70*(3), 539-543.

Bayer, A. E. (1982). A bibliometric analysis of marriage and family literature. *Journal of Marriage and the Family, 44*(3), 527-538.

Bayer, A. E., & Smart, J. C. (1991). Career publication patterns and collaborative "styles" in American academic science. *Journal of Higher Education, 62*(6), 613-636.

Bearinger, L. H. (2006). Beyond objective and balanced: Writing constructive manuscript reviews. *Research in Nursing and Health, 29*(2), 71-73.

Beatty, S. E., Bandyopadhyay, S., Chae, M.-S., & Tarasingh, P. S. (1982). A closer look at manuscript reviewing in marketing. *Journal of Marketing Education, 14*(3), 3-14.

Beaver, M., Gottlieb, N., Rosenblatt, A. (1983). Dilemmas in Manuscript- Evaluation. *Social Work, 28*(4), 326.

Bedeian, A. G. (2003). The manuscript review process: The proper roles of authors, referees, and editors. *Journal of Management Inquiry, 12*(4), 331-338.

Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of management learning and education, 3*(2), 198-216.

Bence, V., & Oppenheim, C. (2004). The role of academic journal publications in the UK Research Assessment Exercise. *Learning Publishing, 17*(1), 53-68.

Benda, W. G. G., & Engels, T. C. E. (2010). The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. *International Journal of Forecasting*, doi:10.1016/j.ijforecast.2010.05.012.

Benedek, E. P. (1976). Editorial practices of psychiatric and related journals: Implications for women. *American Journal of Psychiatry, 133*(1), 89-92.

Bero, L. A., Glantz, S. A., & Rennie, D. (1994). Publication Bias and public health policy on environmental tobacco smoke. *Journal of the American Medical Association, 272*(2), 133-136.

Beyer, J. M. (1978). Editorial policies and practices among leading journals in four scientific fields. *The Sociological Quarterly, 19*(1), 68-88.

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*(5), 1219-1260.

Biddle, C., & Aker, J. (1996). How does the peer review process influence AANA Journal article readability? *Journal of the American Association of Nurse Anaesthetists, 64*(1), 65-68.

Bingham, C. (1999). Peer review on the Internet: Are there faster, fairer, more effective methods of peer review? In F. Godlee, & T. Jefferson (Eds.), *Peer Review in Health Sciences (*pp. 205-223). London: Blackwell BMJ Books.

Bingham, C. M., Higgins, G., Coleman, R., Van Der Weyde, M. B. (1998). The Medical Journal of Australia internet peer-review study. *The Lancet, 352*(9126), 441-445.

Black, N., van Rooyen, S., Godlee, F., Smith, R., Evans, S. (1998). What makes a good reviewer and a good review for a general medical journal? *Journal of the American Medical Association, 280*(3), 231-233.

Blackburn, J. L., & Hakel, M. D. (2006). An examination of sources of peer-review bias. *Psychological Science, 17*(5), 378-382.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *The American Economic Review, 81*(5), 1041-1067.

Bligh, J. (1998). What happens to manuscripts submitted to the journal? *Medical Education, 32*, 567-570.

Bohannon, R. W. (1986). Agreement among reviewers. *Physical Therapy, 66*(9), 1431-1432.

Boice, R., Barlow, D. H., Johnson, K., & Klosko, J. (1984). Behaviourists as peer reviewers: Do they misbehave? *Behaviour Therapist, 27,* 105-107.

Bonjean, C. M., & Hullum, J. (1978). Reasons for journal rejection: An analysis of 600 manuscripts. *Political Science, 11*(4), 480-483.

Borry, P., Schotsmans, P., & Dierickx, K. (2006). Field report; author, contributor or just a signer? A quantitative analysis of authorship trends in the field of bioethics. *Scholarly Publishing, 20*(4), 213.

Brackbill, Y., & Korten, F. (1975). Journal reviewing practices: Authors' and APA members' suggestions for revision. *American Psychologist, 30*(9), 937-940.

Bradley, J. V. (1981). Pernicious publication practices. *Bulletin of the Psychonomic Society, 18*(1), 31-34.

Braun, T., & Bujdoso, E. (1983). Gate-keeping patterns in the publication of analytical chemistry research. *Talanta, 30*(3), 161-167.

Bridgstock, M.W. (1991). The Quality of Single and Multiple Authored Papers - An Unresolved Problem. *Scientometrics, 21*(1), 37-48.

Broad, W. J. (1981). The publishing game: Getting more for less: Meet the Least Publishable Unit, one way of squeezing more papers out of a research project. *Science, 211*(4525), 1137-1139.

Brown, R. J. C. (2007). Double anonymity in peer review within the chemistry periodicals community. *Learning Publishing, 20*(2), 131-137.

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research, 1*(1), 8-22.

Budd, J. (1988). Publication in library and information science: The state of the literature. *Library Journal, 113*(14), 125-131.

Buffardi, L. C., & Nichols, J. A. (1981). Citation impact, acceptance rate, and APA journals. *American Psychologist, 36*(11), 1453-1456.

Burman, K. D. (1982). Hanging from the masthead: Reflections on authorship. *Annals of Internet Medicine, 97*, 602-605.

Burnham, J. C. (1990). The evolution of editorial peer review. *Journal of American Medical Association, 263*(10), 1323-1329.

Cain, J. (1999). Why Be My Colleague's Keeper? Moral Justifications for Peer Review. *Science and Engineering Ethics, 5,* 531-540.

Callaham, M. L., & Tercier, J. (2007a). Qualitative profile of journal peer reviewers and predictors of peer reviewer quality. *PloS Medicine, 4*(1), e40.

Callaham, M. L., & Tercier, J. (2007b). The relationship of previous training and experience of journal peer reviewers to subsequent review quality. *PloS Medicine, 4*(1), 32-40.

Callaham, M. L., Baxt, W. G., Waeckerle, J. F., & Wears, R. L. (1998). Reliability of editors' subjective quality ratings of peer reviewers of manuscripts. *Journal of American Medical Association, 280*(3), 229-231.

Callaham, M. L., Knopp, R. K., & Gallagher, E. J. (2002). Effect of written feedback by editors on quality of reviewers: Two randomized trials. *Journal of American Medical Association, 287*(21), 2781-2783.

Callaham, M. L., Wears, R. L., & Waeckerle, J. F. (1998). Effect of attendance at a training session on peer reviewer quality and performance. *Annals of Emergency Medicine, 32*(3), 318-322.

Calnan, M., Smith, G. D., Sterne, J. A. C. (2006). The publication process itself was the major cause of publication bias in genetic epidemiology. *Journal of Clinical Epidemiology, 59*(12), 1312-1318.

Camp, R. C. (1989). *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, WI: ASQC Quality Press.

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

Campanario, J. M. (1996a). Have referees rejected some of the most-cited articles of all times? *Journal of the American Society for Information Science, 47(*4), 302-310.

Campanario, J. M. (1996b). The competition for journal space among referees, editors, and other authors and its influence on journals' impact factors. *Journal of the American Society for Information Science, 47*(3), 184-192.

Campion, M. A. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personal Psychology, 46*(3), 705-718.

Carnegie, M. E. (1975). The Editor's Report-1975. *Nursing Research, 24*, 243.

Ceci, S. J., & Peters, D. (1984). How blind is blind review? *American Psychologist, 39*(12), 1491-1494.

Chang, J.-J., & Lai, C.-C. (2001). Is it worthwhile to pay referees? *Southern Economic Journal, 68*(2), 457-463.

Chew, F. S. (1988). Coauthorship in radiology journals. *American Journal of Radiology, 150*, 23-26.

Chew, F. S. (1991). Fate of manuscripts rejected for publication in the AJR. *American Journal of Roentgenology, 156*, 627-632.

Chew, M., van der Weyden, M. & Villanueva, E. V. (2005). More than a decade in the life of the impact factor. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Cho, M. K., & Bero, L. A. (1994). Instruments for assessing the quality of drug studies published in the medical literature. *Journal of American Medical Association, 272*(2), 101-104.

Cho, M. K., Justice, A. C., Winker, M. A., Berlin, J. A., Waeckerie, J. F., Callaham, M. L., & Rennie, D. (1998). Masking author identity in peer review: What factors influence masking success? *Journal of American Medical Association, 280*(3), 243-245.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless science: Peer review and U.S. science policy*. Albany, NY: State University of New York Press.

Cicchetti, D. V. (1980). Reliability of reviews for the American Psychologist: A biostatistical assessment of the data. *American Psychologist, 35*(3), 300-303.

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

Cicchetti, D. V., & Eron, L. D. (1979). The reliability of manuscript reviewing for the Journal of Abnormal Psychology. *Proceedings of the American Statistical Association, 22*, 596-600.

Cleary, J. D., & Alexander, B. (1988). Blind versus non-blind review: Survey of selected medical journals. *Drug Intelligence & Clinical Pharmacy, 22*(7), 601-602.

Cobo, E., Selva-O'Callagham, A., Ribera, J.-M., Cardellach, F., Dominguez, R., & Vilardell, M. (2007). Statistical reviewers improve reporting in biomedical articles: A randomized trial. *PloS One, 3*, 1-8.

Coe, R. K. (1984). Evaluating the management journals: a second look. *Academy of Management Journal, 27*(3), 660-666.

Coe, R. K., & Weinstock, I. (1967). Editorial policies of major economic journals. *Quarterly Review of Economics and Business, 7*(4), 37-43.

Coe, R. K., & Weinstock, T. (1968). Publication policies in major business periodicals. *Southern Journal of Business, 3*(1), 1-10.

Coe, R. K., & Weinstock, I. (1983). Evaluating the finance journals: The department chairperson's perspective. *Journal of Financial Research, 6*(4), 345-349.

Coe, R. K., & Weinstock, I. (1984). Evaluating the management journals: A second look. *Academy of Management Journal, 27*(3), 660-666.

Coelho, R.J., & La Forge, J. (2000). Manuscript characteristics affecting reviewers' decisions for rehabilitation counselling related journals. *Journal of Rehabilitation, 66*(2), 4-8.

Cofer, C. N. (1985). Some reactions to manuscript review from a questionnaire study. *The Behavioural and Brain Sciences, 8,* 745-746.

Colaianni, L. A. (1994). Peer review in journals indexed in Index Medicus. *Journal of American Medical Association, 272*(2), 156-158.

Cole, S., Simon, G., & Cole, J. R. (1988). Do journal rejection rates index consensus? *American Sociological Review, 53*(1), 152-156.

Coleman, A. S. (2006). Assessing the value of a journal beyond the impact factor: Journal of Education for Library and Information Science. *Journal of the American Society for Information Science and Technology, 58*(8), 1148-1161.

Cook, T. D. (1985). Postpositivist critical multiplism. In L. Shortland & M. M. Mark (Eds.), *Social science and social policy*, pp. 21-62. Beverly Hill, CA: Sage.

Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research and Practice, 17*(2), 136-137.

Cowen, E. L., Spinell, A., Hightower, A. D., & Lotyczewski, B. S. (1987). Author reaction to the manuscript revision process. *American Psychologist, 42*(4), 403-405.

Crandall, R. (1978). Interrater agreement on manuscripts is not so bad! *American Psychologist, 33*(6), 623-624.

Crane, D. (1967). The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *American Sociologist, 1*(4), 195-201.

Creswell, J. W. (2002). *Qualitative and quantitative approaches.* Upper Saddle River, New Jersey: Sage.

Cronin, B., & McKenzie, G. (1992). The trajectory of rejection. *Journal of Documentation, 48*(3), 310-317.

Cullen, D. J., & Macaulay, A. (1992). Consistency between peer reviewers for a clinical specialty journal. *Academic Medicine, 67*(12), 856-859.

Cullen, J., Joyce, J., Hassall, T., & Broadbent, M. (2003). Quality in higher education: From monitoring to management. *Quality Assurance in Education, 11*(1), 5-14.

Cummings, L. L., Frost, P. J., & Vakil, T. F. (1985). The manuscript review process: A view from the inside on coaches, critics and special cases. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 469-508). Irwin, IL: Homewood.

Cunningham, S. J., & Dillon, S. M. (1997). Authorship patterns in information systems. *Scientometrics, 39*(1), 19-27.

Dancik, B. P. (1991). The importance of peer review. *Serials Librarian, 19*(3/4), 91-94.

Daniel, H.-D. (1993a). An evaluation of the peer review process at *Angewandte Chemie*. *Angewandte Chemie, 32*(2), 234-238.

Daniel, H.-D. (1993b). *Guardians of Science: Fairness and reliability of peer review.* New York: Weiheim.

Dardik, H (1977). Multiple authorship. *Surgery, Gynecology & Obstetrics, 145*, 418.

Davidoff, F., DeAngelis, C. D., Drazen, J. M., Nicholls, M. G., Hoey, J., et al. (2001). Sponsorship, authorship and accountability. *Journal of American Medical Association, 165*(6), 786-788.

Dehmer, P. (1982). APS reviewers refereeing procedures. *Physics Today, 35*(2), 9, 95-97.

DeMaria, A. N. (2007). How do I get a paper accepted? *Journal of the American College of Cardiology, 49(*15), 1666-1667.

Denscombe, M. (2008) Communities of practice: A research paradigm for the mixed methods approach. *Journal of Mixed Methods Research, 2*(3), 270-283.

Denzin, N. K., & Lincoln, Y. S. (Eds.) (2000). *Handbook of qualitative research.* Thousand Oaks, California: Sage Publications.

Department of Education, Training and Youth Affairs. (2000). *The Australian higher education quality assurance framework.* Canberra: Department of Education, Training and Youth Affairs.

DeVaney, T. A. (2001). Statistical significance, effect size, and replication: What do the journals say? *The Journal of Experimental Education, 69*(3), 310-320.

Diamond, D. (1969). Multi-authorship explosion. *New England Journal of Medicine, 280*(1969), 1484-1485.

Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., & Smith, H. (1987). *Publication bias and clinical trials. Controlled Clinical Trials, 8*(4), 343-353.

Dixon, G. F., Schonfeld, S. A., Altman, M., & Whitcomb, M. E. (1983). The peer review and editorial process: A limited evaluation. *The American Journal of Medicine, 74*(3), 494-495.

Donmoyer, R. (1996). Educational research in an era of paradigm proliferation: What's a journal editor to do? *Educational Researcher, 25*(2), 19-25.

Drenth, J. P. H. (1998). Multiple authorship: The contribution of senior authors. *Journal of American Medical Association, 280*(3), 219-221.

Earnshaw, J. J., Farndon, J. R., Guillou, P. J., Johnson, C. D., Murie, J. A., & Murray, G. D. (2000). A comparison of reports from referees chosen by authors or journal editors in the peer review process. *Annals of the Royal College of Surgeons of England, 82*(4), 133-135.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867-872.

Eastwood, S., Derish, P., Leash, E., & Ordway, S. (1996). Ethical issues in biomedical research: perceptions and practices of postdoctoral research fellows responding to a survey. *Science and Engineering Ethics, 2*(1), 89-114.

Eberley, S., & Warner, W. K. (1990). Fields or subfields of knowledge: Rejection rates and reviewer agreement in peer review. *The American Sociologist, 21*(3), 217-231.

Edwards, L. N., & Ferber, M. A. (1986). Journal reviewing practices and the progress of women in the economics profession: Is there a relationship? *Newsletter of the Committee on the Status of Women in the Economics Profession, 1-7*.

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*(8), 1011-1012.

Eisenhart, M. (2002). The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education, 32*(2), 241-255.

Eldredge, J. D. (1997). Identifying peer-reviewing journals in clinical medicine. *Bulletin of the Medical Library Association, 85*(4), 418-422.

Ellison, G. (2002). The slowdown of the economics publishing process. *The Journal of Political Economy, 110*(5), 947-993.

Ely, J., Wooley, M., Lynch, F., McDonald, J., Findlay, L., Choi, Y., & Woolley, K. (2005). Factors affecting the time from manuscript submission to manuscript acceptance. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Emden, C. (1996). Manuscript reviewing: Too long a concealed form of scholarship? *Nursing Inquiry, 3*(4), 195-199.

Engers, M., & Gans, J. S. (1998). Why referees are not paid (enough)? *The American Economic Review, 88*(4), 1341-1349.

Enserink, M. (2001). Peer review and quality: A dubious connection? *Science, 293*(5538), 2187-2188.

Epstein, R. J. (1993). Six Authors in Search of a Citation: Villains or Victims of the Vancouver Convention? *British Medical Journal, 306*, 765-767.

Epstein, W. M. (1990). Confirmational response bias among social work journals. *Science, Technology, & Human Values, 15*(1), 9-38.

Epstein, W. M. (2004). Confirmational response bias and the quality of the editorial processes among American social work journals. *Research on Social Work Practice, 14*(6), 450-458.

Ernst, E., Resch, K. L., Uher, E. M. (1992). Reviewer bias. *Annals of Internal Medicine, 116*(11), 958.

Ernst, E., Saradeth, T., & Resch, K. L. (1993). Drawbacks of peer review. *Nature, 363,* 296.

Estrada, C., Kalet, A., Smith, W., & Chin, M. H. (2006). How to be an outstanding reviewer for the Journal of General Internal Medicine and other journals? *Journal of General Internal Medicine, 21*(3), 281-283.

Evans, A. T., McNutt, R. A., Fletcher, S. W., & Fletcher, R. H. (1993). Characteristics of peer reviewers who produce good quality reviews. *Journal of General Internal Medicine, 8*(5), 422-428.

Fabiato, E. (1994). Anonymity of reviewers. *Cardiovascular Research, 28*(8), 1134-1139.

Fagan, W. T. (1990). To accept or reject: Peer review. *Journal of Educational Thought, 24*(2), 103-113.

Fagley, N. S., & McKinney, I. J. (1983). Reviewer bias for statistically significant results: A re-examination. *Journal of Counselling Psychology, 30*(2), 296-300.

Federico, J., Lee, M., Boozer, C. H., & Diaz, D. (1984). Report of a survey comparing sixty-nine journals in the dental profession. *Educational Directions in Dental Hygiene, 9*, 9-18.

Feinstein, A. R. (1991). Construction, consent, and condemnation in research on peer review. *Journal of Clinical Epidemiology, 44*(4/5), 339-341.

Feinstein, A. R., & Spitzer, W. O. (1989). The peer review process - and an acknowledge of our peerless reviewers. *Journal of Clinical Epidemiology, 42*(1), 1-4.

Ferber, M. A., & Teiman, M. (1980). Are women economists at a disadvantage in publishing journal articles? *Eastern Economic Journal, 6*(3-4), 189-193.

Feurer, I. D., Becker, G. J., Picus, D., Ramirez, E., Darcy, M. D., & Hicks, M. E. (1994). Evaluating peer review. *Journal of American Medical Association, 272*(2), 98-100.

Finn, C. E. (1986). Strengths (and weaknesses) of peer review. *Educational Researcher, 15*(7), 14-15.

Finn, C. E. (2002). The limits of peer review. *Education Week, 21*(34), 30, 34.

Fisher, M., Friedman, S. B., & Strauss, B. (1994). The effect of blinding on acceptance of research papers by peer review. *Journal of American Medical Association, 272*(2), 143-146.

Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist, 45*(5), 591-598.

Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Revalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Journal of American Medical Association, 280(*3), 222-224.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions,* 2nd ed. New York: John Wiley.

Fletcher, R. H., & Fletcher, S. W. (1999). The effectiveness of journal peer review. In F. Godlee, & T. Jefferson (Eds.), *Peer Review in Health Sciences* (pp. 62-75). London: Blackwell BMJ Books.

Flick, U. (2002). *An introduction to qualitative research*, 2nd ed. London: Sage.

Flowerdew, J. (2001). Attitudes of journal editors to non-native speaker contribution. *TESOL Quarterly, 35*(1), 121-149.

Fogg, L., & Fiske, D. W. (1993). Foretelling the judgments of reviewers and editors. *American psychologist, 48*(3), 293-294.

Fölster, S. (1995). The perils of peer view in economics and other sciences. *Journal of Evolutionary Economics, 5*(1), 43-57.

Fontaine, S. I. (1995). With writers' eyes: Perception and change in manuscript review procedures. *Rhetoric Review, 13*(2), 259-264.

Forbes, I. (2003). Perceptions of cutting edge research in UK social science. *Innovation, 16*(3), 271-291.

Fox, M. F. (1994). Scientific misconduct and editorial and peer review process. *The Journal of Higher Education, 65*(3), 298-309.

Frank, E. (1996). Editors' requests of peer reviewers: A study and a proposal. *Preventive Medicine, 25*(2), 102-104.

Frantz, T. T. (1968). Criteria for publishable manuscripts. *Personnel and Guidance Journal, 47*, 384-386.

Freda, M. C., & Kearney, M. (2005). An international survey of nurse editors' roles and practices. *Journal of Nursing Scholarships, 37*(1), 87-94.

Freese, L. (1979). On changing some role relationships in the editorial review process. *The American Sociologist, 14*(4), 231-246.

Friesinger, G. C. (1986). Who should be an author? *Journal of the American College of Cardiology, 8*, 1240-1242.

Fyfe, J. J. (1994). Cops and Robbers in Academe: Editing *Justice Quarterly*. In Simon, R. J., & Fyfe, J. J. (Eds.), *Editors as gatekeepers: Getting published in the social sciences* (pp. 59-70). Boston: Rowman and Littlefield.

Gallagher, E. B., & J. Ferrante. (1991). Agreement among peer reviewers for a middle-sized biomedical journal. In *Peer Review in Scientific Publishing: Papers from the First International Congress on Peer Review in Biomedical Publication* (pp. 153-158). Council of Biology Editors, Chicago, IL.

Gans, J. S., & Shepherd, G. B. (1994). How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives, 8*(1), 165-179.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 115-126.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of American Medical Association, 295*(1), 90-93.

Garfunkel, J. M., Lawson, E. E., Hamrick, H. J., & Ulshen, M. H. (1990a). Effect of acceptance or rejection on the author's evaluation of peer review of medical manuscripts. *Journal of American Medical Association, 263*(10), 1376-1378.

Garfunkel, J. M., Ulshen, M. H., Hamrick, H. J., & Lawson, E. E. (1990b). Problems identified by secondary review of accepted manuscripts. *Journal of American Medical Association, 263*(10), 1369-1371.

Garfunkel, J. M., Ulshen, M. H., Hamrick, H. J., & Lawson, E. E. (1994). Effects of institutional prestige on reviewers' recommendations and editorial decisions. *Journal of American Medical Association, 272*(2), 137-138.

Garlick, S., & Pryor, G. (2004). *Benchmarking the university: Learning about improvement – A report for the Department of Education, Science and Training*. Canberra: DEST.

Garvey, W. D., Lin, N., & Tomita, K. (1972). Research studies in patterns of scientific communication: III. Information-exchange processes associated with the production of journal articles. *Information Storage and Retrieval, 8*(5), 207-221.

Ghali, W. A., Cornuz, J., McAlister, F. A., Wasserfallen, J.-B., Devereaux, P. J., & Naylor, C. N. (2002). Accelerated publication versus usual publication in 2 leading medical journals. *Canadian Medical Association Journal, 166*(9), 1137-1143.

Gibbons, J. D. (1990). U.S. institutional representations on editorial boards of U.S. statistics journals. *American Statistician, 44*(3), 210-213.

Gidez, L. I. (1991). The peer review process: Strengths and weaknesses - A survey of attitudes, perceptions, and expectations. *Serial Librarian, 19*(3/4), 75-85.

Gilbert, J. R., Williams, E. S., & Lungberg, G. D. (1994). Is there gender bias in JAMA's peer review process? *Journal of American Medical Association, 272*(2), 139-142.

Gilliland, S. W., & Beckstein, B. A. (1996). Procedural and distributive justice in the editorial revieww process. *Personnel Psychology, 49*(3), 669-691.

Gilliland, S. W., & Cortina, J. M. (1997). Reviewer and editor decision making in the journal review process. *Personnel Psychology, 50*(2), 427-452.

Ginsberg, L. (1999). Reviewers, orthodoxy, and the passion to publish. *Research on Social Work Practice, 9*(1), 100-103.

Gleditsch, N. P. (2002). Double-blind but more transparent. *Journal of Peace Research, 39*(3), 259-262.

Glen, W. (1989). Musings on the review process. *PALAIOS, 4*(4), 397-399.

Glenn, N. D. (1979). Accountability and the journal article review process. *Contemporary Sociology, 8(*6), 785-787.

Glidewell, J. C. (1988). Reflections on thirteen years of editing AJCP. *American Journal of Community Psychology, 16*, 759-770.

Glogoff, S. (1988). Reviewing the gatekeepers: A survey of referees of library journals. *Journal of the American Society for Information Science, 39*(6), 400-407.

Godlee, F. (2002). Making reviewers visible: Openness, accountability, and credit. *Journal of American Medical Association, 287*(21), 2762-2765.

Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. *Journal of American Medical Association, 280*(3), 237-240.

Golde, C. M., & Walker, G. E. (Eds.) (2006). *Envisioning the Future of Doctoral Education: Preparing stewards of the discipline - Carnegie essays on the doctorate*. Jossey-Bass.

Goldsmith, L. A., Blalock, E., Bobkova, H., & Hall, R. P. (2005). Effect of authors suggestions concerning reviewers on manuscript acceptance. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Goldwyn, R. M. (2005). Do we reject too little and publish too much? *Learned Publishing, 18*(4), 243.

Gomes, S., & Meadows, J. (1998). Perceptions of electronic journals in British universities. *Journal of Scholarly Publishing, 29*(3), 174-181.

Goodman, N. W. (1994). Survey of fulfilment of criteria for authorship in published medical research. *British Medical Journal, 309*(6967), 1482-1485.

Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Annals of Internal Medicine, 121*, 11-21.

Gordon, M. (1979). Peer review in physics. *Physics Bulletin, 30*, 112-113.

Gordon, M. (1980). The role of referees in scientific communication. In J. Hartley (Ed.), *The Psychology of Written Communication: Selected readings* (pp. 263-275). London: Kogan Page.

Gordon, M. D. (1984). How authors select journals: A test of the reward maximization model of submission behaviour. *Social Studies of Science, 14*(1), 27-43.

Gore, S. M., Jones, I. G., & Rytter, E. C. (1977). Misuse of statistical methods: Critical assessment of articles in BMJ from January to March 1976. *British Medical Journal, 1*, 85-87.

Gosden, H. (2003). 'Why not give us the full story?' Functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes, 2*(2), 87-101.

Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist, 33*(10), 920-934.

Graham, G. (1997). The editor: the bridge from authorship to production. *Learned Publishing, 10*(1), 60-63.

Grant, L., & Fine, G. A. (1992). Sociology unleashed: Creative directions in classic ethnography. In M.D. LeCompte, W.L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 405-446). San Diego: Academic Press, Inc.

Graue, M. E. & Grant, C. (2005). Thinking collaboratively about the peer-review process for journal-article publication. *Harvard Educational Review 75*(3), 268-271.

Graue, M. E. (2006). The Transformative Power of Reviewing. *Educational Researcher, 35(*9), 36-41.

Green, S. M., & Callaham, M. L. (2006). Current status of peer review at Annals of Emergency Medicine. *Annals of Emergency Medicine, 48*(3), 304-308.

Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research, 2*(1), 7-22.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*, 255–274.

Groves, T. (2010). Best practice in peer review and editing, ensuring article quality. *British Medical Journal, 13*, 6-8.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review, 69*(2), 190-212.

Haas, L., Milton, S., & Quinn, A. (1996). Surviving the publishing process: A beginners' guide. *RQ, 36*(2), 230-246.

Hackett, E. J. (1997). Peer review in science and science policy. In M. S. Frankel & J. Cave (Eds.), *Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe* (pp. 51-60). Budapest: Central European University Press.

Hamermesh, D.S. (1994). Facts and myths about refereeing. *Journal of Economic Perspectives, 8*(1), 153-163.

Hamermesh, D.S., & Oster, S.M. (2002). Tools or toys? The impact of high technology on scholarly productivity. *Economic Inquiry, 40*(4), 539-555.

Hanna, S. (1996). The editor's role; the editor's vision. *Financial Counselling and Planning, 4*, 1-4.

Harcum, E. R., & Rosen, E. F. (1993). *The Gatekeepers of Psychology: Evaluation of peer review by case history*. Westport: Praeger.

Hargens, L. L. (1988). Scholarly consensus and journal rejection rates. *American Sociological Review, 53*(1), 139-151.

Hargens, L. L., & Herting, J. R. (1990). A new approach to referees' assessments of manuscripts. *Social Science Research, 19*(1), 1-16.

Hargens, L. L. & J. R. Herting. (2006). Analysing the Association between Referees' Recommendations and Editors' Decisions. *Scientometrics, 67*, 15-26.

Harman, G. (1998). Supporting quality research in institutions of higher education. *Australian Journal of Education, 42*(3), 285-302.

Harman, G., & Meek, V.L. (2000). *Reposition quality: Assurance and accreditation in Australian higher education*. Canberra: DETYA.

Harter, S. P. (1998). Scholarly communication and electronic journals: an impact study. *Journal of the American Society for Information Science, 49*(6), 507-516.

Harvey, L., & Green, D. (1993). 'Defining quality'. *Assessment & Evaluation in Higher Education, 18*(1), 9-35.

Hatch, C. L., & Goodman, S. N. (1998). Perceived value of providing peer reviewers with abstracts and preprints of related published and unpublished papers. *Journal of American Medical Association, 280*(3), 273-274.

Hayden, G. F., & Saulsbury, F. T. (1982). A review of the *Journal of Paediatrics*: the first 50 years. *Journal of Paediatrics, 101*, 5-11.

Henderson, A. (2001). Undermining peer review. *Society, 38*(2), 47-54.

Hendrick, C. (1976). Editorial comment. *Personality and Social Psychology Bulletin, 2*, 207-208.

Hendrick, C. (1977). Editorial comment. *Personality and Social Psychology Bulletin, 3,* 1-2.

Henson, K. T. (1988). Writing for education journals. *Phi Delta Kappa, 69*(10), 752-754.

Henson, K. T. (1993). Writing for successful publication: Advice from editors. *Phi Delta Kappan, 74*(10), 799-802.

Hernon, P. P., & Schwartz, C. C. (2003). Peer review. *Library & Information Science Research, 25*, 359-361.

Hernon, P., Smith, A., & Croxen, M. B. (1993). Publication in College & Research Libraries: Accepted, rejected, and published papers, 1980-1991. *College Research Libraries, 54*(4), 303-321.

Hirsch, L. (2004). Randomized clinical trials: What gets published, and when? *Canadian Medical Association Journal, 170*(4), 481-483.

Hoen, W. P., Walvoort, H. C., & Overbeke, A. J. P. M. (1998). What are the factors determining authorship and the other of the authors' names? *Journal of American Medical Association, 280*(3), 217-218.

Hoey, J. (2004). Conflicts of interests and investments. *Canadian Medical Association Journal, 171*(11), 1313.

Hopps, J. G. (1989). Peer review: A trust, not a vault. *Social Work, 34*(1), 3-4.

Houlihan, D., Hofschulte, L., Sachau, D., & Patten, C. (1992). Critiquing the peer review process: Examining a potential dual role conflict. *American Psychologist, 47*(2), 1679-1681.

Howard, L., & Wilkinson, G. (1999). Peer review and editorial decision-making. *Neuroendocrinology Letters, 20*(5), 256-260.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives, 10*(3), 153-158.

Huxley, P (1986). Statistical errors in papers in the British Journal of Social Work (volumes 1-14). *British Journal of Social Work, 16*(6), 645-658.

Ilakovac, V., Fister, K., Marušić, M., & Marušić, A. (2007). Reliability of disclosure forms of authors' contribution. *Canadian Medical Association Journal, 176*(1), 41-46.

Ingelfinger, F. J. (1974). Peer review in biomedical publication. *The American Journal of Medicine, 56*(5), 686-692.

Jackson, N. (2001). Benchmarking in UK HE: An overview. *Quality Assurance in Education, 9*(4), 218-235.

Jacobsen, E. (1980). Authors, reviewers, and editors. *Psychopharmacology, 71*, 111-115.

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*(1), 157-173.

Jauch, L. R., & Gentry, J. W. (1976). Perceptions of faculty evaluation in the soft sciences: A case study. *Research in Higher Education, 5*(2), 159-170.

Jayasinghe, U. W., Marsh, H., & Bond, N. (2001). Peer review in the funding of research in higher education: The Australian experience. *Educational Evaluation and Policy Analysis, 23*(4), 343-364.

Jean-Pierre, E. N., Walvoort, H. C., & Overbeke, A. J. P. M. (1996). Readers' evaluation of effect of peer review and editing on quality of articles in the Nederlands Tijdschrift voor Geneeskunde. *The Lancet, 348*(9040), 1480-1483.

Jefferson, T., Smith, R., Yee, Y., Drummond, M., Pratt, M., & Gale, R. (1998). Evaluating the *BMJ* guidelines for economic submission: Prospective audit of economic submissions to *BMJ* and *the Lancet*. *Journal of the American Medical Association, 280*(3), 275-277.

Johnson, B., & Christensen, L. (2002). *Educational research: Quantitative, qualitative, and mixed approaches, 2nd ed*. Boston: Pearson Education, Inc.

Johnson, R. B., Onwuegbuzie, A.J., Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Johnston, S. C., Lowenstein, D. H., Ferriero, D. M., Messing, R. O., Oksenberg, J. R., & Hauser, S. L. (2007). Early editorial manuscript screening versus obligate peer review: A randomized trial. *Annals of Neurology, 61*, A10-A12.

Juhasz, S., Calvert, E., Jackson, T., Kronick, D. A., & Shipman, J. (1975). Acceptance and rejection of manuscripts. *IEEE Transactions on Professional Communication, PC-18*(3), 177-185.

Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., & Rennie, D. (1998). Does masking author identity improve peer review quality? A randomized controlled trial. *Journal of American Medical Association, 280*(3), 240-242.

Justice, A. C., Berlin, J. A., Fletcher, S. W., Fletcher, R. H., & Goodman, S. N. (1994). Do readers and peer reviewers agree on manuscript quality? *Journal of American Medical Association, 272*(2), 117-119.

Kahn, A. S., Presbury, J. H., Moore, H. B., & Driver, J. D. (1990). Characteristics of accepted versus rejected manuscripts. *Psychology of Women Quarterly, 14*(1), 7-14.

Kassirer, J. P., & Campion, E. W. (1994). Peer review: Crude and understudied, but indispensable. *Journal of the American Medical Association, 272*(1), 96-97.

Katz, D. S., Proto, A. V., & Olmsted,W. W. (2002). Incidence and nature of unblinding by authors: Our experience at two radiology journals with double-blinded peer review policies. *American Journal of Roentgenology, 179*(6), 1415-1417.

Kawano, T., Kehle, T. J., Clark, E., & Jenson, W. R. (1993). School psychology journals: Relationships with related journals and external and internal quality indices. *Journal of School Psychology, 31*(3), 407-424.

Kearney, M. H., & Freda, M. C. (2005). Nurse editors' view on the peer review process. *Research in Nursing & Health, 28*(6), 444-452.

Kerr, S., Tolliver, J., & Petree, D. (1977). Manuscript characteristics which influence acceptance for management and social science journals. *Academy of Management Journal, 20*(1), 132-141.

Kiley, M., & Wisker, G. (2009). Threshold concepts in research education and evidence of threshold crossing. *Higher Education Research and Development, 28*(4), 431-441.

King, C. R., McGuire, D. B., Longman, A. J., & Carroll-Johnson, R. M. (1997). Peer review, authorship, ethics, and conflict of interest. *Image: Journal of Nursing Scholarship, 29*(2), 163-167.

Kirk, S. A. (1993). The puzzles of peer perusal. *Social Work Research & Abstracts, 29*(2), 3-4.

Kirk, S. A., & Franke, T. M. (1997). Agreeing to disagree: A study of reliability of manuscript reviews. *Social Work Research, 21*(2), 121-126.

Kliewer, M. A., Freed, K. S., DeLong, D. M., Pickhardt, P. J., & Provenzale, J. M. (2005). Reviewing the reviewers: Comparison of review quality and reviewer characteristics at the American Journal of Roentgenology. *American Journal of Roentgenology, 184*(6), 1731-1735.

Knoll, E. (1990). The communities of scientists and journal peer review. *Journal of the American Medical Association, 263*(10), 1330-1332.

Knox, F. G. (1981). No unanimity about anonymity. *The Journal of Laboratory and clinical medicine, 97*(1), 1-3.

Koch-Weser, D., & Yankauer, A. (1993). The authorship and fate of international health papers submitted to the American Journal of Public Health in 1989. *American Journal of Public Health, 83*(11), 1618-1620.

Kogan, M. (1986). *Educational accountability*. London: Hutchinson.

Konradsen, J., & Munk-Jorgensen, A. (2007). The destinies of the low- and middle-income country submissions. *ACTA Psychiatrica Scandinavica, 115*(4), 331-334.

Kronick, D. A. (1990). Peer review in 18th-century scientific journalism. *Journal of the American Medical Association, 263*(10), 1321-1322.

La Forge, J., & Coelho, R. J. (1998). Writing for publication: Advice from authors. *Journal of Applied Rehabilitation Counselling, 29*(2), 35-39.

Laband, D. N. (1985). Publishing favouritism: A critique of departmental rankings based on quantitative publishing performance. *Southern Economic Journal, 52*(2), 510-515.

Laband, D. N. (1987). A qualitative test of journal discrimination against women. *Eastern Economic Journal, 8*(2), 149-153.

Laband, D. N. (1990). Is there value-added from the review process in economics? Preliminary evidence from authors. *Quarterly Journal of Economics, 105*(2), 341-352.

Laband, D. N., & Piette, M. J. (1994a). A citation analysis of the impact of blind peer review. *Journal of American Medical Association, 272*(2), 147-149.

Laband, D. N., & Piette, M. J. (1994b). Favouritism versus search for good papers: Empirical evidence regarding the behaviour of journal editors. *Journal of Political Economy, 102*(1), 194-302.

Laband, D. N., & Piette, M. J. (1994c). The relative impacts of economics journals: 1970-1990. *Journal of Economic Literature, 32*(2), 640-666.

Laband, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *The Journal of Political Economy, 108*(3), 632-660.

Lacy, W., & Busch, L. (1982). The institutional and professional context for rural sociology. In: D. Dillman, & D. Hobbs (Eds.), *Rural Society in the US: Issues for the 1980s* (pp. 404-413). Boulder, Colorado: Westview Press.

Laine, C., & Mulrow, C. (2003). Peer review: Integral to science and indispensable to Annals. *Annals of Internal Medicine, 139*(12), 1038-1040.

Lamont, M. (2009). *How Professors Think: Inside the curious world of academic judgment*. Cambridge: Harvard University Press.

Lavelle, J. (1966). Facts of journal publishing, IV. *Publications of the Modern Language Association of America, 81(*6), 3-12

Lazarus, D. (1980). Authors, editors, and referees. *Physical Review Letters, 45*(19), 1527-1528.

Lee, K. P., Boyd, E. A., Holroyd-Leduc, J. M., Bacchetti, P., & Bero, L. A. (2006). Predictors of publication: Characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Medical Journal of Australia, 184*(12), 621-626.

Leimu, R., & Koricheva, J. (2005). What determined the citation frequency of ecological papers? *TRENDS in Ecology and Evolution, 20*(1), 28-32.

Liesegang, T. J., Shakh, M., & Crook, J. E. (2005). The outcome of manuscripts submitted to the American Journal of Ophthalmology between 2002 and 2003. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Linden, W., Craig, K. D., & Wen, F. K. (1992). Contributions of reviewer judgments to editorial decision-making for the Canadian Journal of Behavioural Science: 1985-1986. *Canadian Journal of Behavioural Science, 24*(4), 433-441.

Lindsay, A. (1994). Quality and management in Australian universities. *Journal of Tertiary Education Administration, 16*(1), 55-68.

Lindsey, D. (1976). Distinction, achievement, and editorial board membership. *American Psychologist, 31(*8), 799-804.

Lindsey, D. (1991). Precision in the manuscript review process: Hargens and Herting revisited. *Scientometrics, 22*(2), 313-325.

Lindsey, D. (1992). Improving the quality of social work journals. *Research on Social Work Practice, 2*(4), 515-524.

Lindsey, D. (1999). Ensuring standards in social work research. *Research on Social Work Practice, 9*(1), 115-120.

Link, A. M. (1998). US and non-US submissions: An analysis of reviewer bias. *Journal of American Medical Association, 280*(3), 246-247.

Linkov, F., Lovalekar, M., & LaPorte, R. (2007). Quality control of epidemiological lectures online: Scientific evaluation of peer review. *Croatian Medical Journal, 48*(2), 249-255.

Lloyd, M. E. (1990). Gender factors in reviewer recommendations for manuscript publication. *Journal of Applied Behaviour Analysis, 23*(4), 539-543.

Lock, S. (1985). *A Difficult Balance: Editorial peer review in medicine.* Philadelphia: ISI Press.

Lock, S. (1994). Does editorial peer review work? *Annals, of Internal Medicine, 121*(1), 60-61.

Lock, S., & Smith, J. (1990). What do peer reviewers do? *Journal of American Medical Association, 263*(10), 1341-1343.

Loonen, M. P. J., Hage, J. J., & Kon, M. (2005). Who benefit from peer review? An analysis of the outcome of 100 requests for review by Plastic and Reconstructive Surgery. *Plastic and Reconstructive Surgery, 116*(5), 1461-1472.

Lovat, T., Holbrook, A., & Burke, S. (2008). Ways of knowing in doctoral examination: How well is the doctoral regime? *Educational Research Review, 3*, 66-76.

Lovitts, B. E., (2007). *Making the Implicit Explicit: Creating performance expectations for the dissertation.* Sterling, Virginia: Stylus Publishing.

Lu, Y. (2006). Editorial practices in education: Mapping the field. *New Zealand Journal of Educational Studies, 41*, 7-25.

Lund, H., & Jackson, N. (2000). Creating the conditions for benchmarking: A brief history of Performance assessment in UK HE. In N. Jackson & H. Lund (Eds.), *Benchmarking for higher education* (pp. 13-25). Buckingham: Society for Research into Higher Education & Open University Press.

Mackie, C. D. (1998). *Canonizing economic theory.* Armond, NY: M.E. Sharpe.

Macnab, N., & Thomas, G. (2007). Quality in research and the significance of community assessment and peer review: Education's idiosyncrasy. *International Journal of Research and Method in Education, 30*(3), 339-352.

Madden, A. D. (2000). When did peer review become anonymous? *Aslib Proceedings, 52*(8), 273-276.

Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. *Journal of Advertising, 24*(4), 78-87.

Mahoney, M. J. (1976). *The scientist as subject: The psychological imperative.* Cambridge: Ballinger

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*(2), 161-175.

Manheim, F. T. (1973). Referees and the publications crisis. *EOS, 54*(2), 532-537.

Manheim, F. T. (1975). The scientific referee. *IEEE Transactions on Professional Communication, PC-18*(3), 190-195.

Marginson, S., & van der Wende, M. (2007). *Globalisation and Higher Education.* Education Working Paper No 8, OECD.

Mark Ware Consulting. (2008). *Peer review in scholarly journals: Perspective of the scholarly community – an international study.* Bristol, UK: Author.

Marsh, H. W., & Ball, S. (1981). Inter-judgmental reliability of reviews for the Journal of Educational Psychology. *Journal of Educational Psychology, 73*(6), 872-880.

Marsh, H. W., & Ball, S. (1989). The peer review process used to evaluate manuscript submitted to academic journals: Inter-judgmental reliability. *Journal of Experimental Education, 57*, 150-169.

Marshall, E. (1995). Peer review: Written and unwritten rules. *Science, 270*(5244), 1913.

Martin, A. L. (2003). *AUQA 2002 institutional audit reports: Analysis and comment*. Melbourne: AUQA.

Marušić, A., Bates, T., Anic, A., & Marušić, M. (2006). How the structure of contribution disclosure statements affects validity of authorship: A randomized study in a general medical journal. *Current Medical Research and Opinion, 22*(6), 1035-1044.

Mason, P. M., Steagall, J. W., & Fabritius, M. M. (1992). Publication delays in articles in economics what to do about them. *Applied Economics, 24*(8), 859-874.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, California: Sage.

McCartney, J. L. (1973a). Manuscript Reviewing. *The Sociological Quarterly, 14*(3), 450, 440-446.

McCartney, J. L. (1973b). Selecting reviewers. *The Sociological Quarterly, 14*(2), 287-288, 290.

McCarty, R. (2002). Science, politics, and peer review: An editor's dilemma. *American Psychologist, 57*(3), 198-201.

McCloskey, J. C. (1977). Publishing opportunities for nurses: A comparison of 65 journals. *Nurse Educator, 2*(4), 4-13.

McCloskey, J. C., & Swanson, E. (1982). Publishing opportunities for nurses: A comparison of 100 journals. *Journal of Nursing Scholarship, 14*(2), 50-56.

McCutchen, C. W. (1991). Peer review: Treacherous servant, disastrous master. *Technology Review, 94*(7), 28-37.

McDowell, J. M., & Melvin, M. (1983). The determinants of co authorship: An analysis of the economics literature. *The Review of Economics and Statistics, 65*(1), 155-160.

McKinnon, K. R., Walker, S. H., & Davis, D. (2000). *Benchmarking: A manual for Australian universities*. Canberra: DETYA, Higher Education Division.

McNamee, S. J., & Willis, C. L. (1994). Stratification in science: A comparison of publication patterns in four disciplines. *Science Communication, 15*(4), 396-416.

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 American Medical Association, 263*(10), 1371-1376.

McPeek, M. A., DeAngelis, D. L., Shaw, R. G., Moore, A. J., Rausher, M. D., Strong, D. R., Ellison, A. M., Barrett, L., Rieseberg, L., & Breed, M. D. (2009). The Golden Rule of Reviewing. *The American Naturalist, 173*(5), E155-E158.

McReynolds, P. (1971). Reliability of ratings of research papers. *American Psychologist, 26*(4), 400-401.

Medoff, M. H. (2003). Editorial favouritism in economics? *Southern Economic Journal, 70*(2), 425-434.

Menon, M. E. (2003). Views of teaching-focused and research-focused academics on the mission of higher education. *Quality in Higher Education, 8*(1), 39-54.

Mertens, D. M. (2005). *Research and evaluation in education and psychology. Integrating diversity with quantitative, qualitative, and mixed methods,* (2nd ed.). Thousand Oaks: CA: Sage Publication

Metoyer-Duran, C. (1993). The readability of published, accepted, and rejected papers appearing in College & Research Libraries. *College & Research Libraries, 54*(6), 517-525.

Meyer, C. H. (1983). Responsibility in publishing. *Social Work, 46*(1), 3.

Meyer, D. (2000). On the information and selection function of economics articles - claims and reality. *Journal of Post Keynesian Economics, 23*(1), 49-74.

Miller, A. C., & Serzan, S. L. (1984). Criteria for identifying a refereed journal. *Journal of Higher Education, 55*(6), 673-699.

Miller, J., & Perrucci, R. (2001). Back stage at Social Problems: An analysis of the editorial decision process, 1993-1996. *Social Problems, 48*(1), 93-110.

Misakian, A. L., & Bero, L. A. (1998). Publication bias and research on passive smoking. *Journal of American Medical Association, 280*(3), 250-253.

Miser, H. J. (1998). The easy chair: Journal editing as I see it. *Interfaces, 28*(5), 115-123.

Moher, D., Dulberg, C. S., & Wells, G. A. (1994). Statistical power, sample size, and their reporting in randomized controlled trials. *Journal of American Medical Association, 272*(2), 122-124.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*(1), 48.

Morrison, G. H. (1985). Positive peer pressure. *Analytical Chemistry, 57*(8), 1505.

Morrow, J. R., Bray, M. S., Fulton, J. E., & Thomas, J. R. (1992). Interrater reliability of 1987-1991 Research Quarterly for Exercise and Sport Reviews. *Research Quarterly for Exercise and Sport, 63*(2), 200-203.

Morse, J. M. (1996). "Revise and resubmit": Responding to reviewers' reports. *Qualitative Health Research, 6*(2), 149-151.

Moss, P. A., Phillips, D. C., Erickson, F. D., Floden, R. E., Lather, P. A., & Schneider, B. L. (2009). Learning from Our Differences: A dialogue across Perspectives on Quality in Education Research. *Educational Researchers, 38*(7), 501-517.

Moyer, R. C., & Crockett, J. H. (1976). Academic journals: Policies, trends and issues. *Academy of Management Journal, 19*(3), 489-495.

Mruck, K., & Mey, G. (2002). Between printed past and digital future. *Research in Science Education, 32*(2), 257-268.

Munley, P. H., Sharkin, B., & Gelso, C. J. (1988). Reviewer ratings and agreement on manuscripts reviewed for the Journal of Counselling Psychology. *Journal of Counselling Psychology, 35*(2), 198-202.

Murphy, P. S. (1995). Benchmarking academic research output in Australia. *Assessment & Evaluation in Higher Education, 20*(1), 45-57.

Murphy, T. M., & Utts, J. M. (1994). A retrospective analysis of peer review at Physiologia Plantarum. *Physiologia Plantarum, 92*(4), 535-542.

Mussurakis, S. (1993). Coauthorship trends in the leading radiological journals. *Acta Radiologica, 34*(4), 316-320.

Mylonas, C. (1992). Research in neurosurgical journals. *British Journal of Neurosurgery, 6*, 41-46.

Neff, B. D., & Olden, J. D. (2006). Is peer review a game of chance? *Bioscience, 56*(4), 333-340.

Nemery, B. (2001). What happens to the manuscripts that have not been accepted for publication in Occupational and Environmental Medicine? *Occupational and Environmental Medicine, 58,* 604-607.

Neuhauser, D. (1997). Peer review and the research commons: A problem of success. *Medical Care, 35*(4), 301-302.

Neuliep, J. W., & Crandell, R. (1990). Editorial bias against replication research. *Journal of Social Behaviour and Personality, 5*(4), 85-90.

Newman, S. H. (1986). Improving the evaluation of submitted manuscripts. *American Psychologists, 21*(10), 980-981.

Newton, J. (2002). Views from below: Academics coping with quality. *Quality in Higher Education, 8*(1), 39-61.

Nicholas, D., Huntington, P., & Rowlands, I. (2005a). Open access journal publishing: The views of some of the world's senior authors. *Journal of Documentation, 61*(4), 497-519.

Nicholas, D., Jamali, H. R., Huntington, P., & Rowlands, I. (2005b). In their very own words: Authors and scholarly journal publishing. *Learned Publishing, 18*(3), 212-220.

Nicholas, D., Jamali, H. R., Huntington, P., & Rowlands, I. (2006). On the tips of their tongues: Authors and their views on scholarly publishing. *Learned Publishing, 19*(3), 193-203.

Noble, K. A. (1989). Publish or Perish: What 23 journal editors have to say. *Studies in Higher Education, 14*(1), 97-102.

Nylenna, M., Riis, P., & Karlsson, Y. (1994). Multiple Blinded reviews of the same two manuscripts: Effects of referee characteristics and publication language. *Journal of American Medical Association, 272*(2), 149-151.

O'Cathain, A., Murphy, E., & Nicholl, J. (2007). Why, and how, mixed methods research is undertaken in health services research: A mixed methods study. *Journal of Mixed Methods Research, 1*(2), 147-163.

O'Connor, D., & Van Orden, P. (1978). Getting into Print. *College and Research Libraries, 39*(5), 391.

Olson, C. M., Rennie, D., Cook, D., Dickersin, K., Flanagin, A., Hogan, J. W., Zhu, Q., Reiling, J., & Pace, B. (2002). Publication bias in editorial decision making. *Journal of American Medical Association, 287*(21), 2825-2828.

Onwude, J.L., Staines, A., & Lilford, R. J. (1993). Multiple author trend worst in medicine. *British Medical Journal, 306*, 1345.

Opthof, T., Furstner, F., Van Geer, M., & Coronel, R. (2000). Regrets or no regrets? No regrets! The fate of rejected manuscripts. *Cardiovascular Research, 45*(1), 255-258.

Opthof, T., Coronel, R., & Janse, M. J. (2002a). Submissions, impact factor, reviewer's recommendations and geographical bias within the peer review system (1997-2002): Focus on Germany. *Cardiovascular Research, 55*, 215-219.

Opthof, T., Coronel, R., & Janse, M. J. (2002b). The significance of the peer review process against the background of bias: Priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovascular Research, 56*, 339-346.

Osborne, M. D., & Brady, D. J. (2002). The room is long and narrow. *Research in Science Education, 32*(2), 163-170.

Over, R. (1994). Use of peer review by the Australian Research Council. *Australian Universities' Review, 37*(1), 31-35.

Overbeke, J., & Wager, E. (1999). The state of evidence: What we know and what we don't know about journal peer review. In F. Godlee, & T. Jefferson (Eds.), *Peer Review in Health Sciences* (pp. 45-61). London: Blackwell BMJ Books.

Oxman, A. D., Guyatt, G. H., Singer, J., Goldsmith, C. H., Hutchison, B. G., Milner, R. A., & Streiner, D. L. (1991). Agreement among reviewers of review articles. *Journal of Clinical Epidemiology, 44*(1), 91-98.

Pardeck, J. T. (1992). Are social work journal editorial boards competent? Some disquieting data with implications for Research on Social Work Practice. *Research on Social Work Practice, 2*(4), 487-496.

Parker, G. (1986). On blinding the journal assessor. *Australian and New Zealand Journal of Psychiatry, 20*(2), 241-242.

Parker, G., Barnett, B., Holmes, S., & Manicavasagar, V. (1984). Publishing in the parish. *Australian & New Zealand Journal of Psychiatry, 18*, 78-85.

Patterson, S. C. (1994). The itch to publish in political science. In R. J. Simon & J. J. Fyfe (Eds.), *Editors as Gatekeepers: Getting published in the social sciences.* Maryland: Rowman & Littlefield Publishers.

Patton, M. Q. (2002). *Qualitative research and evaluation methods. 3rd ed.* Thousand Oaks, California: Sage.

Perakyla, A. (2005). Analysing talk and text. In N.K. Denzin & Y.S. Lincoln (Eds.), *The Sage handbook of qualitative research, 3rd ed.*, (pp. 69-886). Thousand Oaks, California: Sage.

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioural and Brain Sciences, 5*, 187-255.

Pfeffer, J., Leong, A., & Strehl, K. (1977). Paradigm development and particularism: Journal publication in three scientific disciplines. *Social Forces, 55*(4), 938-951.

Plug, C. (1993). The reliability of manuscript evaluation for the South African Journal of Psychology. *South African Journal of Psychology, 23*(1), 43-48.

Polak, J. F. (1995). The role of the manuscript reviewer in the peer review process. *American Journal of Roentgenology, 165*, 685-688.

Posey, L. M. (2005). Behaviours of authors and peer reviewers following change from closed to open system. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science, 10*(1), 95-101.

Price, J. H., & Dake, J. A. (2002). Ethical guidelines for manuscript reviewers and journal editors. *American Journal of Health Education, 33*(4), 194-197.

Psathas, G. (1976). Replies to Calvin Peters. *The American Sociologist, 11*(3), 173-174.

Purcell, G. P., Donovan, S. L., & Davidoff, F. (1998). Changes to manuscripts during the editorial process: Characterizing the evolution of a clinical paper. *Journal of American Medical Association, 280*(3), 227-228.

Ray, J., Berkwits, M., & Davidoff, F. (2000). The fate of manuscripts rejected by a general medical journal. *The American Journal of Medicine, 109*(2), 131-135.

Rehr, H., Showers, N., Rosenberg, G., Blumenfield, S. (1998). Profiles of published social work practitioners: Who wrote and why. *Social work in Health Care, 28*(2), 83-94.

Reindl, T. (2004). State funding for higher education: The squeeze continues. *College and University, 79*(4), 35-36.

Relman, A. S. (1978). Are we a filter or a sponge? *The New England Journal of Medicine, 299*(4), 197.

Relman, A. S. (1990). Peer review in scientific journals - What good is it? *Western Journal of Medicine, 153*(5), 520-522.

Relman, A. S., & Angell, M. (1989). How good is peer review? *The New England Journal of Medicine, 321*(12), 827-829.

Rennie, D. (1993). More peering into editorial peer review. *Journal of the American Medical Association, 270*(23), 2856-2858.

Rennie, D. (1998). Freedom and responsibility in medical publication. *Journal of the American Medical Association, 280*(3), 300-302.

Rennie, D. (1999). Editorial peer review: Its development and rationale. In F. Godlee, & T. Jefferson (Eds.), *Peer Review in Health Sciences* (pp. 1-13). London: Blackwell BMJ Books.

Resch, K. I., Ernst, E., & Garrow, J. (2000). A randomized controlled study of reviewer bias against an unconventional therapy. *Journal of the Royal Society of Medicine, 93*, 164-167.

Richardson Jr., J. V. (2002). The peer review process: Acceptances, revisions, and outright rejections (1). *Library Quarterly, 72*(1), v-xi.

Roberts, J. C., Fletcher, R. H., & Fletcher, S. W. (1994). Effects of peer review and editing on the readability of articles published in Annals of Internal Medicine. *Journal of American Medical Association, 272*(2), 119-121.

Robertson, P. (1976). Towards open refereeing. *New Scientist, 71*, 410.

Robin, E. D., & Burke, C. M. (1987). Peer review in medical journals. *Chest, 91*(2), 252-255.

Robinson, D. H., & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5), 21-27.

Rodman, H., & Mancini, J. A. (1977). Editors, manuscripts, and equal treatment. *Research in Higher Education, 7*(4), 369-374, 377.

Roediger III, H. L. (1987). The role of journal editors in the scientific process. In D. N. Jackson, & J. P. Rushton (Eds.), *Scientific Excellence: Origins and assessment* (pp. 222-252). Newbury Park, CA: Sage.

Rohrich, R. J. (2006). The promise and perils of peer review. *Plastic and Reconstructive Surgery, 118*(3), 795-797.

Rojewski, J. W., & Domenico, D. M. (2004). The art and politics of peer review. *Journal of Career and Technical Education, 20*(2), 41-53.

Rosenblatt, A., & Kirk, S. A. (1980). Recognition of authors in blind review of manuscripts. *Journal of Social Service Research, 3*(4), 383-394

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., Hachinski, V. C., Gibbons, R. J., Gardner, T. J., & Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *Journal of American Medical Association, 295*(14), 1675-1680.

Rotton, J., Foos, P., Van Meek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behaviour and Personality, 10*, 1-13.

Rotton, J., Levitt, M., & Foos, P. (1993). Citation impact, rejection rates, and journal value. *American Psychologist, 48*(8), 911-912.

Rowlands, I., & Nicholas, D. (2006). The changing scholarly communication landscape: An international survey of senior researchers. *Learned Publishing, 19*(1), 31-55.

Rowney, J. A., & Zenisek, T. J. (1980). Manuscript characteristics influencing reviewers' decisions. *Canadian Psychology, 21*(1), 17-20.

Rubin, H. R., Redelmeier, D. A., Wu, A. W., & Steinberg, E. P. (1993). How reliable is peer review of scientific abstracts? *Journal of General Internal Medicine, 8*(5), 255-258.

Runeson, G., & Loosemore, M. (1999). Gate-keepers or judges: Peer reviews in construction management. *Construction Management and Economics, 17*(4), 529-539.

Sandelowski, M. (1995). Triangles and crystals: On the geometry of qualitative research. *Research in Nursing & Health, 18(*6), 569-574.

Satyanarayana, K., & Ratnakar, K. V. (1989). Authorship patterns in life sciences preclinical basic and clinical research papers. *Scientometrics, 17*(3-4), 363-371.

Scarr, S., & Weber, B. L. R. (1978). The reliability of reviews for the American Psychologist. *American Psychologist, 33*, 935.

Schaeffer, D. L. (1970). Do APA journals play professional favourites? *American Psychologist, 25*, 362-365.

Scharschmidt, B. F., DeAmicis, A., Bacchetti, P., & Held, M. J. (1994). Chance, concurrence, and clustering: Analysis of reviewers' recommendations on 1000 submissions to The Journal of Clinical Investigation. *The Journal of Clinical Investigation, 93*(5), 1877-1880.

Schauder, D. (1994). Electronic publishing of professional articles: Attitudes of academics and implications for the scholarly communication industry. *Journal of the American Society for Information Science, 45*(2), 73-100.

Scherer, R. W., Dickersin, K., & Langenberg, P. (1994). Full publication of results initially presented in abstracts. *Journal of American Medical Association, 272*(2), 158-162.

Schriger, D. L., Arora, S., Schriger, V. A., & Altman, D. G. (2005). An analysis of the content of medical journals' instructions for authors. *Presentation to the Fifth International Congress on Peer Review and Biomedical Publication, September, 16-18, 2005, Chicago, Illinois*. Retrieved on 25 July 2006 from http://www.ama-assn.org/public/peer/program.html.

Schriger, D. L., Sinha, R., Schroter, S., Liu, P. Y., & Altman, D. G. (2006). From submission to publication: A retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Annals of Emergency Medicine, 48*(6), 750-756.

Schroter, S., & Groves, T. (2004). BMJ training for peer reviewers. *British Medical Journal, 328*, 658.

Schroter, S., Black, N., Evans, S., Carpenter, J., Godlee, F., & Smith, R. (2004). Effects of training on quality of peer review: Randomised controlled trial. *British Medical Journal,* doi:10.1136/bmj.38023.700775.AE (published 2 March 2004).

Schroter, S., Tite, L., Hutchings, A., & Black, N. (2006). Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. *Journal of American Medical Association, 295*(3), 314-317.

Scott, A. (2007). Peer review and the relevance of science. *Futures, 39*(7), 827-845.

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

Seligman, L. (1986). The manuscript evaluation process used by AACD journals. *Journal of Counselling and Development, 65,* 189-192.

Sellers, S. L., Mathiesen, S. G., Smith, T., Perry, R. (2006). Perceptions of professional social work journals: Findings from a national survey. *Journal of Social Work Education, 42*(1), 139-159.

Shapiro, D. W., Wenger, N. S., & Shapiro, M. F. (1994). The contributions of authors to multi-authored biomedical research papers. *Journal of American Medical Association, 271*(6), 438-442.

Sharp, D. W. (1990). What can and should be done to reduce publication bias. *Journal of American Medical Association, 263*(10), 1390-1391.

Shaw, D. V. (1994). To publish or not? *SRA Journal, 26*(2), 5-8.

Sherrell, D. L., Hair, J. F., & Griffin, M. (1989). Marketing academicians' perceptions of ethical research and publishing behaviour. *Journal of the Academy of Marketing Science, 17*(4), 315-324.

Siegelman, S. S. (1989). Radiology statement of purpose 1990. *Radiology, 173*, 467-468.

Siegelman, S. S. (1991). Assassins and zealots: Variations in peer review. *Radiology, 178*, 637-642.

Sigelman, L., & Whicker, M. L. (1987). Some implications of bias in peer review: A simulation-based analysis. *Social Science Quarterly, 68*(3), 494-509.

Silverman, D. (2006). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction* (3rd edition). London: Sage.

Silverman, R. J. (1976). The education journal editor: A portrait. *Journal of Education, 158*(4), 39-68.

Simon, R. J., Bakanic, V., & McPhail, C. (1986). Who complains to journal editors and what happens. *Sociological Inquiry, 56*(2), 259-271.

Simon, R. J., & Fyfe, J. J. (Eds.) (1994). *Editors as gatekeepers: Getting published in the social sciences.* Boston: Rowman and Littlefield.

Smart, J. C., & Bayer, A. E. (1986). Author collaboration and impact: A note on citation rates of single and multiple authored articles. *Scientometrics, 10*, 297-305.

Smigel, E. O., & Ross, H. L. (1970). Factors in the Editorial Decision. *The American Sociologist, 5*(1), 19-21.

Smith, K. J., & Dombrowski, R. F. (1998). An examination of the relationship between author-editor connections and subsequent citations of auditing research articles. *Journal of Accounting Education, 16*(3/4), 497-506.

Smith, K. J., & Laband, D. N. (1995). The role of editor's professional connections in determining which papers get published: Evidence from accounting research journals. *Accounting Perspectives, 1*, 21-30

Smith, M. F. (2003). Is higher education worth funding? *Academe, 89*(2), 109-110.

Smith, R. (1994). Promoting research into peer review. *British Medical Journal, 309*, 143-144.

Smith, R. (1997). Peer review: Reform or revolution? *British Medical Journal, 315*, 759-760.

Smith, R. (1999a). Opening up BMJ peer review: A beginning that should lead to complete transparency. *British Medical Journal, 318,* 4-5.

Smith, R. (1999b). The future of peer review. In F. Godlee, & T. Jefferson (Eds.), *Peer Review in Health Sciences* (pp. 244-253). London: Blackwell BMJ Books.

Smith, R. (2006a). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.

Smith, R. (2006b). The trouble with medical journals. *Journal of the Royal Society of Medicine, 99*(3), 115-119.

Sobal, J., & Ferentz, K. S. (1990). Abstract creep and author inflation. *The New England Journal of Medicine, 323*, 488-489.

Speck, R. L. (1993). *Publication Peer Review: An Annotated Bibliography.* Westport, Connecticut: Greenwood Press.

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology, 20*(8), 357-358.

Squires, B. P. (1987). Reviewing and editing: A shared responsibility. *Canadian Medical Association Journal, 137*(7), 16.

Squires, B. P. (1988). Clarifying requirements for manuscripts. *Canadian Medical Association Journal, 138*(2), 305-306.

Squires, B. P. (1989a). Biomedical manuscripts: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 141*(7), 17-19.

Squires, B. P. (1989b). Biomedical review articles: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 141*(8), 195-197.

Squires, B. P. (1989c). Case reports: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 141*(9), 379-380.

Squires, B. P. (1989d). Descriptive studies: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 141*(11), 879-880.

Squires, B. P. (1989e). Editorials and platform articles: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 141*(10), 666-667.

Squires, B. P. (1990f). Illustrative material: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 142*(5), 447-449.

Squires, B. P. (1990g). Statistics in biomedical manuscripts: What editors want from authors and peer reviewers? *Canadian Medical Association Journal, 142*(3), 213-214.

Squires, B. P. (1991). The role of the editor in peer review. *Serial Librarian, 19*(3/4), 87-89.

Starbuck, W. H. (2003). Turning lemons into lemonade: Where is the value in peer review? *Journal of Management Inquiry, 12*(4), 344-351.

Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science, 16*(2), 180-200.

Staw, B. M. (1985). Repairs on the road to relevance and rigor: Some unexplained issues in publishing organizational research. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the Organizational Sciences* (pp. 97-107). Irwin, IL: Homewood.

Stephenson, S. L. (2004). Saving quality from quality assurance. *Perspectives, 8*(3), 62-67.

Stern, J. M., & Simes, R. J. (1997). Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal, 315*(7109), 640-645.

Stern, P. N. (1996). Putting a manuscript to bed: Operationalizing the process. *Health Care for Women International, 17*(3), v-vii.

Sternberg, R. J., & Gordeeva, T. (1996). The anatomy of impact: What makes an article influential? *Psychological Science, 7*(2), 69-75.

Strayhorn, J., McDermott, J. F., & Tanguay, P. (1993). An intervention to improve the reliability of manuscript reviews for the Journal of the American Academy of Child and Adolescent Psychiatry. *American Journal of Psychiatry, 150,* 947-952.

Street, M. D., Bozeman, D. P., & Whitfield, J. M. (1998). Author perception of positive and negative editor behaviors in the manuscript review process. *Journal of Social Behaviour & Personality, 13*(1), 1-22.

Strub, R. L., & Black, F. W. (1976). Multiple authorship. *Lancet, 2*(7994), 1090-1091.

Stull, G. R. (1989). Peer-review process is key to quality publication. *Ceramic Bulletin, 68*(4), 850-852.

Suls, J., & Martin, R. (2009). The Air We Breathe: A critical look at practices and alternatives in the peer-review process. *Perspectives on Psychological Science, 4*(40), 40-50.

Susser, M., & Yankauer, A. (1993). Prior, duplicate, repetitive, fragmented, and redundant publication and editorial decisions. *American Journal of Public Health, 83*(6), 792-793.

Sutherland, H. J., Meslin, E. M., Da Cunha, R., & Till, J. E. (1993). Judging clinical research questions: What criteria are used? *Social Science in Medicine, 37*(12), 1427-1430.

Swan, A. (1999). What authors want? The ALPSP research study on the motivations and concerns of contributors to learned journals. *Learned Publishing, 12*(3), 170-172.

Swanson, E. A., McCloskey, J. C., & Bodensteiner, A. (1991). Publishing opportunities for nurses: A comparison of 92 U.S. journals. *IMAGE: Journal of Nursing Scholarship, 23*(1), 33-38.

Sweitzer, B. J., & Cullen, D. J. (1994). How well does a journal's peer review process function? A survey of authors' opinion. *Journal of American Medical Association, 272*(2), 152-153.

Tashakkori, A. & Creswell, J. W. (2007a). The new era of mixed methods. *Journal of Mixed Methods Research, 1*(1), 3.

Tashakkori, A. & Creswell, J.W. (2007b). Exploring the nature of research questions in mixed methods research. *Journal of Mixed Methods Research 1*(4), 207-211.

Tashakkori, A., & Teddlie, C. (Eds.) (2003). *Handbook of mixed methods in social & behavioural research*. Thousand Oaks, California: Sage Publications.

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology with Examples. *Journal of Mixed Methods Research, 1*(1), 77-100.

Tenopir, C. (2004). Online Databases-Online Scholarly Journals: How Many? *Library Journal, 129,* 32.

Terry, J. L. (1996). Authorship in College & Research Libraries Revisited: Gender, institutional affiliation, collaboration. *College and Research Libraries, 57*(4), 377-379.

Thyer, B. A., & Myers, L. L. (2003). An empirical evaluation of the editorial practices of social work journals. *Journal of Social Work Education, 39*(1), 125-140.

Tight, M. (2003). Reviewing the reviewers. *Quality in Higher Education, 9*(3), 295-303.

Tite, L., & Schroter, S. (2007). Why do peer reviewers decline to review? A survey. *Journal of Epidemiology and Community Health, 61*, 9-12.

Tobin, K. (2002). The multiple faces of peer review in science education. *Research in Science Education, 32,* 135-156.

Trafford, V. (2003). Questions in doctoral vivas: views from the inside. *Quality Assurance in Education, 11*(2), 114-122.

van Rooyen, S. (2001). The evaluation of peer-review quality. *Learned Publishing, 14*(2), 85-91.

van Rooyen, S., Black, N., & Godlee, F. (1999). Development of the Review Quality Instrument (RQI) for assessing peer reviews of manuscripts. *Journal of Clinical Epidemiology, 52*(7), 625-629.

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1998a). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomized trial. *British Medical Journal, 318*, 23-27.

van Rooyen, S., Godlee, F., Evans, S., Smith, R., & Black, N. (1998b). Effect of blinding and unmasking on the quality of peer review: A randomized trial. *Journal of American Medical Association, 280*(3), 234-237.

Via, B. J. (1996). Publishing in the journal literature of library and information science: A survey of manuscript review processes and acceptance. *College and Research Libraries, 57*(4), 365-376.

Vidovich, L. (2002). Acceding to audits: New quality assurance policy as a 'settlement' in fostering international markets for Australian higher education. *In AARE 2002 conference papers. Conference of the Australian Association for Research in Education, 1 - 5 December 2002, Brisbane.*

Von Glinow, M. A., & Novelli Jr, L. (1982). Ethical standards within organizational behaviour. *Academy of Management Journal, 25*(2), 417-436.

Wackerle, J. F., & Callaham, M. L. (1996). Medical journals and the science of peer review: Raising the standard. *Annals of Emergency Medicine, 28*(1), 75-77.

Wager, E., & Jefferson, T. (2001). Shortcomings of peer review in biomedical journals. *Learned Publishing, 14*(4), 257-263.

Wager, E., Parkin, E. C., & Tamber, P. S. (2006). Are reviewers suggested by authors as good as those chosen by editors? Results of a rater-blinded, retrospective study. *BMC Medicine, 4*(13). Retrieved on 20 November 2007 from http://www.biomedcentral.com/1741-7015/4/13.

Wallace, D. P., & Fleet, C. V. (1998). Qualitative research and the editorial tradition: A mixed metaphor. *Library Trends, 46*(4), 752-768.

Walker, G. E., Golde, C. M., Jones, L., Bueschel, A. C., & Hutchings, P. (2008). *The Formation of Scholars: Rethinking doctorial education for the twenty-first century.* San Francisco, CA: Jossey-Bass.

Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: A randomised controlled trial. *British Journal of Psychiatry, 176*, 47-51.

Warner, W. K., Eberley, S., Johnson, B. L., & Albrecht, S. L. (1985). Discriminants of editorial decision outcomes. *Rural Sociology, 50*(4), 614-625.

Wates, E., & Campbell, R. (2007). Author's version vs. publisher's version: An analysis of the copy-editing function. *Learned Publishing, 20*(2), 121-129.

Watkins, M. W. (1979). Chance and interrater agreement on manuscripts. *American Psychologist, 34*, 796-797.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive Measure: Nonreactive measures in the social sciences.* Chicago: Rand McNally.

Webb, W. B. (1979). Continuing education: Refereeing journal articles. *Teaching of Psychology, 6*(1), 59-60.

Weber, E. J., Callaham, M. L., Wears, R. L., Barton, C., & Young, G. (1998). Unpublished research from a medical specialty meeting. *Journal of American Medical Association, 280*(3), 257-259.

Weber, E. J., Katz, P.P., Waeckerie, J. F., & Callaham, M. L. (2002). Author perception of peer review: Impact of review quality and acceptance on satisfaction. *Journal of American Medical Association, 287*(21), 2790-2793.

Weber, R. P., & Stevenson, W.C. (1981). Evaluations of accounting journals and department quality. *Accounting Review, 56*(3), 596-612.

Weeks, P. (2000). Benchmarking in higher education: An Australian case study. *Innovations in Education and Training International, 37*(1), 59-67.

Weir, E., & Joseph, K. S. (1999). How quickly does CMAJ evaluate submissions? *Canadian Medical Association Journal, 161*(8), 985.

Weller, A. C. (1987). Editorial policy and the assessment of quality among medical journals. *Bulletin of the Medical Library Association, 75*(4), 310-316.

Weller, A. C. (1990a). Editorial peer review in US medical journals. *Journal of American Medical Association, 263*(10), 1344-1347.

Weller, A. C. (1990b). Editorial peer review: Methodology and data collection. *Bulletin of the Medical Library Association, 78*(3), 258-270.

Weller, A. C. (1991). Potential bias in editorial peer review: A study of U. S. medical journals. *Serials Librarian, 19*(3/4), 95-103.

Weller, A. C. (1995). Editorial peer review: Research, current practices, and implications for librarians. *Serials Review, 21*(1), 53-65.

Weller, A. C. (1996). Editorial peer review: A comparison of authors publishing in two groups of U.S. medical journals. *Bulletin of the Medical Library Association, 84*(3), 359-366.

Weller, A. C. (2000). Editorial peer review for electronic journals: Current issues and emerging models. *Journal of the American Society for Information Science, 51*(14), 1328-1333.

Weller, A. C. (2001). *Editorial Peer Review: Its strengths & weakness.* Medford, New Jersey: Information Today.

Wessely, S. (1996). What do we know about peer review? *Psychological Medicine, 26,* 883-886.

Whitehurst, G. J. (1983). Interrater agreement for reviews for developmental review. *Developmental Review, 38*, 73-78.

Whitehurst, G. J. (1984). Interrater agreement for journal manuscript reviews. *American Psychologist, 39*, 22-28.

Wilcox, L. J. (1998). Authorship: The coin of the realm, the source of complaints. *Journal of American Medical Association, 280*(3), 216-217.

Wild, C. J. (1995). Continuous Improvement of Teaching: A Case Study in a Large Statistics Course. *International Statistical Review, 63*(1), pp. 49-68.

Williamson, A. (2003). What will happen to peer review? *Learned Publishing, 16*(1), 15-20.

Willis, C. L., & Bobys, R. S. (1983). Perishing in publishing: An analysis of rejection letters. *Wisconsin Sociologist, 20*(2-3), 84-91.

Wilson, J. D. (1978). Peer review and publication. *Journal of Clinical Investigation, 61,* 1697-1701.

Wilson, J. R. (2002). Responsible authorship and peer review. *Science and Engineering Ethics, 8*(2), 155-174.

Wood, D. J. (1998). Peer review and the Web: The implications of electronic peer review for biomedical authors, referees and learned society publishers. *Journal of Documentation, 54*(2), 173-197.

Wolff, W. M. (1970). A study of criteria for journal manuscripts. *American Psychologist, 25*, 636-639.

Woodhouse, D. (2002). The impact of audit. In *New times new approaches: Conference proceedings, 16th Australian International Education Conference,* Hobart, Tasmania, 30 September – 4 October 2002.

Woodhouse, D. (2003). Quality improvement through quality audit. *Quality in Higher Education, 9*(2), 134-139.

Woolley, C. M. (2009). Meeting the mixed methods challenge of integration in a sociological study of structure and agency. *Journal of Mixed Methods Research, 3*(1), 7-25.

Yamazaki, S. (1995). Refereeing system of 29 life science journals preferred by Japanese scientists. *Scientometrics, 33*(1), 123-129.

Yankauer, A. (1977). Editor's report. *American Journal of Public Health, 67*(2), 136-137.

Yankauer, A. (1979). Editor's report. *American Journal of Public Health, 69*(3), 222-223.

Yankauer, A. (1982). 1983: Editor's report. *American Journal of Public Health, 74*(3), 207-208.

Yankauer, A. (1990). Who are the peer reviewers and how much do they review? *Journal of American Medical Association, 263*(10), 1338-1340.

Yankauer, A. (1991). How blind is blind review? *American Journal of Public Health, 81*(7), 843-845.

Yates, L. (2004). *What does Good Education Research look like?* Maidenhead: Open University Press (Conducting Educational Research Series).

Yin, R. K. (2006). Mixed Methods Research: Are the Methods Genuinely Integrated or. Merely Parallel? *Research in the Schools, 13*(1), 41-47.

Yoels, W. C. (1971). Destiny or dynasty: Doctoral origins and appointment patterns of editors of the American Sociological Review, 1948-1968. *The American Sociologist, 6,* 134-139.

Yoels, W. C. (1974). The structure of scientific fields and the allocation of editorship on scientific journals: Some observations on the politics of knowledge. *The Sociological Quarterly, 15*, 264-276.

Yohe, G. W. (1980). Current publication lags in economics journals. *Journal of Economic Literature, 18,* 1050-1055.

Zairi, M. (1994). Benchmarking: The best tool for measuring competitiveness. *Benchmarking: An International Journal, 1*(1), 11-24.

Zellmer, W. A. (1977). What editors expect of reviewers? *American Journal of Hospital Pharmacy, 34,* 819.

Ziman, J. M. (1968). *Public Knowledge: An essay concerning the social dimension of science*. Cambridge University Press.

Zsindely, S., & Schubert, A. (1989). Editors-in-chief of medical journals: Are they experts, authorities, or neither. *Communication Research, 16*(5), 695-700.

Zsindely, S., Schubert, A., & Braun, T. (1982). Citation patterns of editorial gatekeepers in international chemistry journals. *Scientometrics, 4*(1), 69-76.

Zuckerman, H., & Merton, R.K. (1971). Patterns of evaluation in science: Institutionalization, structure and functions of the referee systems. *Minerva, 9*(1), 66-100.