What We Still Don’t Know About Peer Review

OMAR SABAJ MERUANE, CARLOS GONZÁLEZ VERGARA, and ÁLVARO PINA-STRANGER

Despite criticisms, the peer review process (PRP) is undoubtedly well established as an official and legitimated mechanism for evaluating and controlling scientific production. Although PRP has been a prominent object of study, we argue in this article that empirical research on PRP has not been addressed in a comprehensive way. Nine categories were applied to 150 empirical research articles on the topic with results revealing various gaps in empirical PRP research: (1) the research has been dedicated to the evaluation of the system rather than to the actual description of PRP as a concrete socio-discursive practice; (2) the most productive group of studies considers the multiple relationships between the sociological attributes (socio-demographic or scientometrical) of the actors (authors, reviewers, and editors) and the results of the process but does not take into account the texts exchanged by those actors; and (3) the few studies that do analyze the texts interchanged in the process do not take into account any of the variables included (such as scientometrical data, agreement, and rejection rates) in the more productive areas of the field. This lack of integration among the methodological approaches to PRP results in a partial comprehension of this important process, which determines the production and dissemination of an important part of scientific knowledge.

Keywords: peer review; research articles; production and evaluation of scientific knowledge

Sadly, the majority of peer review research is dilettante science research, practised by scientists who are not trained to observe their own practices.1

In the period 1969–2006, there are 3720 publications under the topic search of the term ‘peer review.’2 Using the ‘advanced search’ function (TS = ‘Peer review’ and Document types = Article) of the Web of Science

(WOS), version 5.10 interface, one can obtain 1241 records, coded as research articles, in the period 2007–13. A similar search on SpringerLink for the period 1969–2013 provides 22,926 publications. In addition, Lutz Bornmann suggests that in order to have a complete picture of the research done on the peer review process (PRP), one should also consider the great amount of grey literature on the topic. In fact, ever since the seminal study of Harriet Zuckerman and Robert Merton, the interest in PRP has never decreased.

There are several reasons for the proliferation of research in this field. First, the importance of PRP to the production, evaluation, and consumption of science is evident. Second, as Bornmann shows, the results of the empirical evidence of the process are inconclusive. Third, PRP has been the subject of multiple criticisms that have encouraged researchers to debate whether the process is biased or unbiased, reliable or unreliable, predictively valid or invalid (or if it is able, or unable, to predict the impact of a paper as measured by the number of citations it generates).

Considering the enormous amount of research on the subject, it would be risky to state that PRP is an under-researched topic. Nevertheless, as a result of the analysis of some of the most-cited reviews on the subject, it can be inferred that the evidence does not tend to be conclusive regarding the multiple types of relationships involved in this process. This is a clear indicator that, despite the great amount of research accumulated, PRP is a field that has yet to be fully defined. In particular, Stefan Hirschauer has pointed out some important gaps, or under-researched areas, in empirical studies of PRP. In this article, we show evidence of these and other gaps in order to promote, as Reinhart does, more interdisciplinary research on the topic.

This article is organized as follows. In the first section, various definitions for PRP are critically revised before a general delineation of the term is given; then, a general review of the topic is provided, describing the main disciplines that have studied PRP and their most frequent objects of study. In the second section, some gaps in the field will be critically described based on the analysis of 150 empirical studies on PRP, using the guidelines of grounded theory. As a result, it will be argued that, despite the great amount of research on the topic, empirical research in the field has not been comprehensively done.
**PRP: Some Preliminary Definitions**

PRP is considered to be the main quality control mechanism of science\(^{11}\) and can be described in the following steps: (1) an author, or a group or authors, sends an article to the editor of a journal; (2) the editor selects a number of reviewers (usually two); (3) the reviewers write a report with evaluative comments and a recommendation (accepted, minor or major revisions, or rejected); and (4) considering the recommendations of the reviewers, the editor makes a decision that is communicated to the author(s).\(^{12}\) More steps can be added, such as when the reviewers are asked to evaluate the changes in the revised manuscript and to state their conformity or disconformity with them. Also, when the referees do not agree in their recommendations, the editor usually selects another reviewer. Several studies report that PRP is implemented with a high degree of heterogeneity among disciplines and journals.\(^{13}\) These differences can be identified in the decision process, the selection of reviewers, or in the criteria used in evaluations. Commonly, PRP is also classified based on the degree of secrecy in the process: double blind, single blind, open, or public.\(^{14}\)

PRP is not easy to define because its characterization strongly depends on the scope used to delineate it. From the point of view of public policies for research, PRP is a scientific quality control mechanism that determines the allocation of resources to finance the scientific industry. For editors, PRP is a system that allows them to ‘separate the wheat from the chaff,’ an input helping them to decide what to publish. For authors, PRP is an obligatory step to have their work published as well as an opportunity to receive feedback from the members of the community, which permits them to improve the quality of their manuscripts. For reviewers, PRP is a way to help maintain the high standards of, and participate in, the generation of knowledge in their fields.

From a socio-economic framework, Flaminio Squazzoni\(^{15}\) considers PRP (and science in general) to be an imperfect economic exchange system because the same actors involved in the system pursue different interests as they assume different roles (authors, reviewers, and editors). Besides, these actors do not share the same information when they take on specific roles due to secrecy requirements. In this article, we use the following general definitions of the term: Conceptually, PRP is a collective action in which actors epistemically coordinate. In particular, it is a socio-technical judgmental discursive practice that determines the
production, dissemination, and consumption of scientific knowledge. Operationally, a single PRP is an interaction among actors who have specific social attributes and exchange texts during the process. Furthermore, the same actor can fulfill different roles (author, reviewer, or editor) in various single PRPs (that is, as an author in one case, as a reviewer in another, and maybe even as an editor in a third case).

**The main disciplines that have studied PRP**
The seminal work of Zuckerman and Merton\(^\text{16}\) shows that the relative status between authors and referees does not influence the rejection or acceptance rates, although status by itself does affect the duration of the process and is correlated with specific decision paths (resulting in verdicts of accepted, rejected, or requiring revision, among others). During the 1960s and 1970s of the last century, PRP was researched, directly or indirectly, by some sociologists of science. The work of the Cole brothers,\(^\text{17}\) heirs of the Mertonian tradition, is a good example of this first stage in the research on PRP. Although the sociology of science was still generating knowledge about PRP,\(^\text{18}\) two disciplines started concentrating the empirical research on PRP in the 1980s: medicine (particularly medical sub-specialties) and psychology. These studies have mainly focused on the evaluation of the system in terms of its reliability, predictive validity, and fairness or the absence (presence) of bias.\(^\text{19}\) All of this research has been conducted mainly by the editors of medical journals who were worried about the criticism of the system. They looked for ways to improve its deficiencies and propose new alternatives.\(^\text{20}\) A summary of these disciplines, their data, and the objects of study is provided in Table 1.

**Methods**
To show the existence of gaps in the empirical research on peer review, three stages were followed: searching and sampling a corpus of empirical research on PRP, cleaning the database, and analyzing the data set. In the first stage, keyword searches were conducted in the most-used scientific databases (such as WOS, Scopus, ScienceDirect, Elsevier, Springer-Link, and JSTOR), using variations of the following terms: scientific communication, peer review, evaluation of science, scientific quality, journals. The references cited by Eugene Garfield,\(^\text{21}\) Bornmann,\(^\text{22}\) and Juan Miguel Campanario\(^\text{23}\) were also used as a point of comparison with our corpus of analysis. The intention was not to produce an exhaustive
corpus but, rather, to have a representative sample of the most prominent areas, journals, and authors that appear in the recognized literature. Concordantly, all articles that did not have full-text availability online were excluded. The great majority of analyzed articles correspond to the study of peer review for journals, although research on the PRP of grant applications was not excluded. Two independent researchers cleaned the database of false positives, which were studies on peer review in other settings (educational and psychological).

The analysis consisted of two steps. First, all analysis was done using the guidelines of grounded theory, which essentially include the general exploration of the data, the generation of relevant categories, the definition and use of those categories to describe all data, and a mechanism for assuring the consistency of the categories in terms of agreement or consensus. After a general exploration of the data, the following categories were generated to classify the corpus in terms of types of studies: (1) empirical studies; (2) essays, editorials, and general information; (3) improvement proposals; (4) reviews (see Table 2). In addition, the doctoral background of each main author was registered (see Table 3).

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Objects of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociology of Science</td>
<td>• Rates of acceptance / rejection / revision</td>
</tr>
<tr>
<td></td>
<td>• Socio-demographic data: sex, age, nationality, affiliation, discipline, and journal</td>
</tr>
<tr>
<td></td>
<td>• Relational data = sociometric, bibliometric, and scientometric data: citation analysis, impact factor, H-index, number of papers published, rankings, prestige, funds obtained, and other indicators.</td>
</tr>
<tr>
<td></td>
<td>• Reward systems (involving prestige and funding)</td>
</tr>
<tr>
<td></td>
<td>• Science production (involving the number of papers published, their quality, and impact)</td>
</tr>
<tr>
<td></td>
<td>• Evaluation patterns</td>
</tr>
<tr>
<td>Medicine / Psychology</td>
<td>• Rates of acceptance / rejection / revision</td>
</tr>
<tr>
<td></td>
<td>• Bias / fairness</td>
</tr>
<tr>
<td></td>
<td>• Agreement of reviewers / reliability</td>
</tr>
<tr>
<td></td>
<td>• Predictive validity</td>
</tr>
<tr>
<td></td>
<td>• Evaluation criteria</td>
</tr>
</tbody>
</table>
The second step of the analysis only considered the articles included in the empirical studies category. Following the same procedure, nine categories were generated to identify the variables, or topics, that were relevant in characterizing empirical research on PRP: (1) rates of acceptance/rejection/revision (minor or major); (2) bias; (3) agreement among reviewers; (4) predictive validity; (5) evaluation criteria; (6) socio-demographic data; (7) sociometric/scientometric data; (8) discourse analysis of isolated texts; and (9) discourse analysis of related texts (see Table 4).

For each category, a question and a description defining that category were proposed along with an example and a counter-example. The three researchers agreed on the pertinence of the examples and the counter-examples, and only the questions and descriptions were edited by consensus. A pair of assistants independently applied, without further
instruction, these nine categories to all of the articles identified as empirical studies, keeping in mind that, as the categories are not exclusive, the same article could respond positively to more than one category. As a means of checking the consistency of the categories, the coders shared one third of the articles to be classified; the resulting percentages of agreement ranged from 60 per cent to 92 per cent. To obtain the final results, all cases in which there was disagreement were resolved by consensus by the three main researchers. Simple frequency and percentages were used to analyze interactions between categories (see Table 5), and

### Table 4. Categories Studied in Empirical Research on PRP

<table>
<thead>
<tr>
<th>Code</th>
<th>Categories</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Rates of acceptance/rejection/revision (minor or major)</td>
<td>38</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td>Bias</td>
<td>46</td>
<td>31</td>
</tr>
<tr>
<td>C</td>
<td>Agreement among reviewers</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>D</td>
<td>Predictive validity</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>E</td>
<td>Evaluation criteria</td>
<td>54</td>
<td>36</td>
</tr>
<tr>
<td>F</td>
<td>Socio-demographic data</td>
<td>124</td>
<td>83</td>
</tr>
<tr>
<td>G</td>
<td>Sociometric/Scientometric data</td>
<td>71</td>
<td>47</td>
</tr>
<tr>
<td>H</td>
<td>Discourse analysis of isolated texts</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>H1</td>
<td>Discourse analysis of related texts</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

### Table 5. Interaction between Categories

<table>
<thead>
<tr>
<th>Code</th>
<th>Categories</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>H1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Rates of acceptance/rejection/revision (minor or major)</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>21</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Bias</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>26</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Agreement among reviewers</td>
<td>6</td>
<td>7</td>
<td>14</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Predictive validity</td>
<td>10</td>
<td>17</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Evaluation criteria</td>
<td>31</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Socio-demographic data</td>
<td>36</td>
<td>7</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Sociometric/Scientometric data</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Discourse analysis of isolated texts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>H1</td>
<td>Discourse analysis of related texts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the number of categories considered in these studies was also computed as an indicator of comprehensiveness in the empirical research on PRP (see Table 6).

## RESULTS

### The Final Data Set

In the first search, 389 articles were collected and analyzed, and fifty-seven were discarded as they were considered to be off-topic. Generally, these corresponded to studies in the field of peer interaction in educational settings. The resulting general data set was then further classified according to the ‘types of study’ categories mentioned earlier. The relevant classes selected are shown in Table 2.

The information in Table 2 partially confirms the disadvantage of empirical research on PRP pointed out by Hirschauer. Although the number of studies analyzed in this article is less than the combined number of references taken from the most-cited reviews in the field (181 references in Campanario; 150 references in Reinhart; and 259 references in Bornmann), it is a significant corpus of references that differs from those in the following aspects: it contains only empirical works, its analysis is made using content categories, and it contains more up-to-date references (thirty-one references from 2011 to 2013).

### The Doctoral Backgrounds of the Main Authors of Empirical Research on PRP

After the revision of the 150 articles on empirical PRP, it was found that the main authors doing empirical research on PRP hold doctorates in different disciplines (see Table 3). As shown in Table 3, our corpus

---

**Table 6. The Number of Categories included in Empirical Studies of PRP**

<table>
<thead>
<tr>
<th>Number of categories included</th>
<th>Number of studies</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–3</td>
<td>111</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>5–7</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Totals</td>
<td>150</td>
<td>100</td>
</tr>
</tbody>
</table>

---
indicates that investigators doing empirical research on PRP are still coming mainly from the fields of medicine/psychology, bibliometrics/scientometrics, and the sociology of science (see Table 1), which support Hirschauer’s and Reinhart’s observations on the lack of integration of other fields in the study of PRP.

**Categories of Empirical Research on PRP**

Table 4 presents the frequencies and percentages for the categories (objects of study) considered in the empirical research on peer review. As Table 4 shows, 83 per cent of the empirical research on PRP includes socio-demographic variables (Category F), which account for the disciplines, journals, countries, age, and/or sex of the actors (authors, reviewers, or editors) under study. Besides this single category, the most prominent group of variables considered in PRP research includes scientometrical variables (Category G), such as prestige, citations, and ranking; the criteria or arguments used in evaluations (Category E); and the presence of bias (Category B). The second most prominent group consists of acceptance/rejection rates (Category A), predictive validity (Category D), and agreement among reviewers (Category C). The last group includes the discourse analysis of isolated texts (Category H) as well as those that are connected (Category H1), both of which are the least frequent types of work in the empirical field of PRP.

Table 5 shows the percentages of interaction between categories (that is, the percentage of studies that share two variables). From Table 5, we can see that the variables that most interact are Categories F and G. In general, Categories H and H1 are the ones that have the least interaction with the rest of the categories. The highest percentage of 36 per cent (Categories F and G) shows that more than a third of empirical studies include both socio-demographical data along with scientometrical variables. In general, what Table 5 indicates is that the empirical study of PRP only addresses particular aspects of the process and that the texts interchanged by the actors in the process are not prominent objects of study in the field.

Table 6 shows the number of categories included in the data of analysis. The majority of the studies (74%) consider one to three categories. Only 15 per cent of the articles analyzed includes four categories, and a minority (11%) considers more than five categories. This is an indirect measure of the comprehensiveness of the empirical research on PRP.
TABLE 7. Categories Grouped into Classes

<table>
<thead>
<tr>
<th>Category/Classes</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (systemic)</td>
<td>A. Rates of acceptance/rejection/modification (major or minor)</td>
</tr>
<tr>
<td></td>
<td>B. Bias</td>
</tr>
<tr>
<td></td>
<td>C. Agreement among reviewers (reliability)</td>
</tr>
<tr>
<td></td>
<td>D. Predictive validity regarding scientific impact/quality</td>
</tr>
<tr>
<td></td>
<td>E. Evaluative criteria/patterns</td>
</tr>
<tr>
<td>Socio-demographic</td>
<td>F. Socio-demographic data: sex, age, nationality, affiliation, discipline, journal</td>
</tr>
<tr>
<td>Relational</td>
<td>G. Sociometric/Scientometric/Bibliometric data: citation analysis, impact factor, H-Index, number of papers published, ranking, reward related (prestige, funding), and so on.</td>
</tr>
<tr>
<td>Content functional</td>
<td>H. Discursive analysis of the content and function (purpose, polarity) of one PRP text, namely, the reviewers’ report.</td>
</tr>
<tr>
<td>Process (and content functional)</td>
<td>H1. Discursive analysis of the content and function (purpose, polarity) of part, or all, of the texts involved in a PRP.</td>
</tr>
<tr>
<td>Multiple role (a non-category class)</td>
<td>The actors in a PRP have had the experience of multiple roles in various PRP’s (in some cases, that of author; in other cases, that of reviewer; and even that of editor)</td>
</tr>
</tbody>
</table>

The Gaps Found in Empirical PRP Research

In Table 7, categories of the same type are placed into classes that will be used to denote the gaps found in the overview of PRP made in this investigation.

In Table 8, the disciplines involved in the study of PRP are also placed into classes based on the objects of study they hold in common. Each group of objects of study is then paired up with the corresponding category classes from Table 7. From there, it can be easily deduced what the gaps are for each class of discipline since they are precisely those category classes that are not covered in their objects of study.

DISCUSSION

The output class of studies (medicine and psychology) has a systemic, not a procedural, conception of peer review. Analyzing typical objects of study in these disciplines (such as agreement, validity, and bias),
one can argue that, in fact, all these topics depend more on the output (recommendation, editorial decision) of the process than on the real, socio-textual, interactions involved in it, resulting in content-functional and process gaps. Georg Steinhause and his colleagues have indicated the inconvenience of measuring peer review with scientific standards.\textsuperscript{35} We concur since, in our view, this field of research (with a focus on the system and not on the process) has been looking for something that is potentially dangerous to scientific development because, as some researchers argue, high agreement among reviewers could be a problem as this might result in a lack of diversity, and even redundancy, in the reviews, whereas disagreement allows for evaluation from a number of different perspectives.\textsuperscript{36} Furthermore, not all bias is necessarily negative as there has been a documented case that favours research that is important, original, well designed, and well reported.\textsuperscript{37}

Another limitation applies specifically to bibliometrics, scientometrics, and science and technology studies exploring peer review (Scientific Indicator 2 in Table 8). These studies are the ones focusing on Category G.

---

**Table 8. The Relationships between Disciplines, Categories, and Gaps**

<table>
<thead>
<tr>
<th>Discipline classes</th>
<th>Disciplines</th>
<th>Objects of study according to categories</th>
<th>Gaps in the objects of study according to category classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (systemic)</td>
<td>Medicine psychology</td>
<td>A, B, C, D, E, F, G</td>
<td>Content-functional process, multiple role</td>
</tr>
<tr>
<td>Scientific indicator</td>
<td>1 Sociology of science</td>
<td>A, D, E, F, G</td>
<td>Output relational</td>
</tr>
<tr>
<td></td>
<td>2 Bibliometrics</td>
<td>G</td>
<td>Relational</td>
</tr>
<tr>
<td></td>
<td>Library and information science</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social network analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sociometrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technology and science studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Discourse Analysis</td>
<td>F, H</td>
<td>Content-functional process, multiple role</td>
</tr>
</tbody>
</table>

---

\textsuperscript{35} Journal of Scholarly Publishing
Citation analysis, impact factor, H-index, and other indicators constitute an independent, established field of empirical research on peer review. Researchers representative of this field not only come from the discipline of library and information science but also from social network analysis. Although this approach has provided multiple, and varied, data that enhances our understanding of the scientific communication process, the debate focuses on the relevance of the indicators as a proxy of scientific quality, or impact, without considering the process itself or the functions (content plus action) that an actor fulfills when interacting with others through a citation. This results in content-functional and process gaps. In fact, only a few works, which are not necessarily in the field of peer review, combine some of the indicators with content, or with functional, analysis categories. The tendency seems to be that if an investigation has to do with indicators, it will not include content, or functional, categories—that is, we can know how many times a paper was cited, but we do not know the purpose, the polarity (positive or negative), or the content of that citation.

Studies dealing with Categories A, D, E, F, and G correspond mainly to the field of the sociology of science (Scientific Indicator 1 in Table 8), which has a long-standing tradition of research on, and concern with, PRP. These studies usually analyze sociological phenomena such as science production, reward systems, and evaluation patterns in PRP. Although more dynamic (procedural) than the medicine and psychology approach, which mainly uses the result of the process as the main variable (basically just Category A), this group rarely includes actual texts (such as articles submitted, reviewers’ reports, and articles published) as data involved in PRP. Instead, surveys and in-depth interviews are the main instruments of data recollection.

To describe the phenomena under study (science production, reward systems, and evaluation patterns), scientists in this area add the sociological attributes of the actors (authors, reviewers, and editors) to the main variable (Category A) used in the fields of medicine and psychology (the result of the process and the editorial decision). These attributes may be socio-demographic (Category F) or scientometrical (Category G). In the first case, actors are classified in terms of a general category, such as sex, age, discipline, or affiliation, among others. Scientometrical attributes are relational in the sense that they represent the relative status, or position, of an actor with respect to the rest of the members of the
community. Papers published, rankings, citations received, prestige, and funds obtained, among others, are typical examples of scientometrical data. In summary, the sociology-of-science approach relates sociological attributes to PRP to describe several phenomena, such as reward systems and editorial decisions, among others, but it does not analyze the texts that are interchanged by the actors throughout the process. All of this results in content-functional and process gaps.

Discourse analysis studies (the content class in Table 8) are relatively scarce in comparison to other fields. Excuses for this lack of research have been given in terms of the confidentiality, or the occluded nature, of the texts involved. As is the case in the empirical research of peer review, ethical issues regarding sensitive data are far from being a simple issue. Discourse analysis studies have two other problems that limit our understanding of PRP. First, and most important, is the fact that scientific discourse researchers tend to ignore advances in other fields, specifically concerning bibliometrical and scientometrical indicators (Category G) as well as the social networks, or mechanisms, underlying scientific knowledge as described by classic works in the sociology of science.

Much of the research using discourse analysis to investigate PRP contains only the socio-demographic data (Category F) of the actors, such as sex, age, discipline, or the journal where they have published, but none of the cases include any relational information, such as papers published and citations received, among other bibliometrical data (Category G). In addition, the editorial decision (Category A), or the result of the process, which is a central variable for the rest of the fields exploring PRP, is commonly excluded in discourse analysis studies. A second restriction of discourse studies on peer review is that they tend to focus on only one specific text of the process, namely the reviewers’ report. This emphasis on only one of the texts impedes analyzing peer review as a process and implies a static, partial account of it. The exception to this last critique is the classic, detailed, and rigorous work of Greg Myers. All of this results in output gaps, relational gaps, and process gaps.

Finally, one additional limitation common to all of the fields described is what Campanario and Cassidy Sugimoto and Blaise Cronin have not sufficiently emphasized: a major flaw of most productive disciplines studying peer review is the independent data treatment for both actors
and the roles they fulfil. A basic issue to be considered when studying peer review is that the same actors can fulfil different roles in various single PRPs. Thus, for example, when analyzing all of the actors (authors, reviewers, and editors) participating in a specific journal, a primary and relevant piece of information to have is which actors are more participative and in which roles, revealing the functional overlaps among actors. This results in a multiple role gap.

As a result, there are various important unresolved questions regarding PRP. Although these questions have to do with the interaction of all of the categories described before, in our view, the most urgent ones are those from the less developed areas—for example, discourse analysis (Categories H and H1). For instance, some of these questions are: Do the discursive characteristics of the revision vary according to the sociometric attributes of the reviewers (Category G-related question)? Is there a discursive structural, or linguistic, difference between articles rejected and accepted (Category A-related question)? Do actors practice what they preach—that is, if you, as a reviewer, usually focus on methods, are your methods sound when acting as an author (multiple role-related question)?

FINAL REMARKS
As we have seen, the main areas in PRP research contain several gaps. Medicine and psychology, the most productive empirical areas in the study of PRP, conceive of it as a system and not as a process, focusing more on the evaluation than on the actual description. The sociology of science and, more recently, library and information science are the second most productive fields. This group, however, pays little attention to the content and functions of the texts interchanged during PRP. Instead, they describe the relationships (using several measures, such as co-authorship, citations, affiliation networks, and so on) among the actors involved in the process. Due to restricted access to data and confidentiality matters, discourse studies are less frequent in the field. Although these studies contain socio-demographic information about the actors, they do not take into account the relationships between the scientometrical attributes of the actors and the discursive actions they use in the texts involved in PRP. In addition, discourse studies describe texts in isolation, mainly the reviewer’s report, without considering that,
in PRP, texts are the products of the interactions of all of the actors who participate in it. Finally, we have noted that, when investigating the actors in PRP, all of these fields treat the sociological attributes separately from the roles the actors fulfill, ignoring the fact that there is usually an overlap—that is, the same actors can participate with different roles in various single PRPs. Figure 1 presents a diagram that shows the gaps in empirical research on PRP in relation to the classes of discipline they apply to.

As these results suggest, the study of PRP requires not only further, but more comprehensive, approaches. We have often found that the ‘black box’ metaphor is used to describe PRP, and as we have argued in this article, despite the massive amount of literature, this box is far from being open.51
OMAR SABAJ is professor of linguistics at Universidad de La Serena, Chile. His main research area is scientific discourse analysis and scientific literacy.

CARLOS GONZÁLEZ VERGARA is professor of Spanish grammar at Pontificia Universidad Católica de Chile. He is also editor-in-chief of Onomázein: Journal of Linguistics, Philology and Translation.

ÁLVARO PINA-STRANGER is professor of socio-economics at Université de Rennes 1 in France. His area of expertise is social network analysis in innovation systems.

ACKNOWLEDGEMENTS
We are especially grateful to Christopher F.S. Maligec, José O. Valderrama, and Paulina Meza for valuable comments on the drafts of this article. The authors thank the support of the Chilean National Fund for Scientific and Technological Development (Fondecyt No. 1130290) and the European Research Council (7th Framework Programme, ERC-2010-StG 263529).

NOTES
5. Bornmann, ‘Scientific Peer Review’
6. Ibid.


11. Bornmann, ‘Scientific Peer Review’


20. Ibid.

21. Garfield, ‘Index of Peer Review HistCite Collections’

22. Bornmann, ‘Scientific Peer Review’

24. Scientometrical attributes are relational in the sense that they represent the relative status or position of an actor with respect to the rest of the members of the community.
26. Hargens, ‘Scholarly Consensus and Journal Rejection Rates’
27. The list of the 150 empirical studies that were analyzed appears in Appendix 1.
28. The final categories, questions, and descriptions used are available from the authors.
30. Campanario, ‘Peer Review for Journals As It Stands Today, Part 2’
31. Reinhart, ‘Peer Review of Grant Applications in Biology and Medicine’
32. Bornmann, ‘Scientific Peer Review’
34. Reinhart, ‘Peer Review of Grant Applications in Biology and Medicine’
36. Campanario, ‘Peer Review for Journals As It Stands Today, Part 1’
37. Bornmann, ‘Scientific Peer Review’
38. Garfield, ‘Index of Peer Review HistCite Collections’; Cassidy Sugimoto and Blaise Cronin, ‘Citation Gamesmanship: Testing for Evidence of Ego Bias in Peer Review,’ *Scientometrics* 95 (2013): 851–62
42. Chubin, ‘Beyond Invisible Colleges’


49. Sugimoto and Cronin, ‘Citation Gamesmanship: Testing for Evidence of Ego Bias in Peer Review’

50. Squazzoni, ‘Peering into Peer Review’

51. Garfield, ‘Index of Peer Review HistCite Collections’
APPENDIX 1: LIST OF REFERENCES ANALYZED


24. Lutz Bornmann and Hans-Dieter Daniel, ‘Selecting Scientific Excellence through Committee Peer Review: A Citation Analysis of Publications Previously Published to Approval or Rejection of Post-Doctoral Research Fellowship Applicants,’ *Scientometrics* 68, 3 (2006): 427–40


73. Timothy Hofer, Steven Bernstein, Sonya DeMonner, and Rodnay Hayward, ‘Discussion between Reviewers Does Not Improve Reliability or Peer Review of Hospital Quality,’ *Medical Care* 38, 2 (2000): 152–61


122. Martin Reinhart, ‘Peer Review of Grant Applications in Biology and Medicine, Reliability, Fairness, and Validity,’ *Scientometrics* 81, 3 (2009): 789–809


129. Mona Shattell, Peggy Chinn, Sandra Thomas, and Richard Cowling, ‘Authors’ and Editors’ Perspectives on Peer Review Quality in Three
What We Still Don’t Know About Peer Review

137. Cassidy Sugimoto and Blaise Cronin, ‘Citation Gamesmanship: Testing for Evidence of Ego Bias in Peer Review,’ Scientometrics 95 (2013): 851–62


148. Paul Wouters, ‘Citation Cycles and Peer Review Articles,’ *Scientometrics* 38, 1 (1997): 39–55
