

# Peering into peer review: Good quality reviews of research articles require neither writing too much nor taking too long

## *Revisando os pareceres de arbitragem: uma avaliação da qualidade de artigos de pesquisa não requer escrever muito nem demorar demais*

Paula CABEZAS Del FIERRO<sup>1</sup>  0000-0003-2257-1907

Omar SABAJ MERUANE<sup>1</sup>  0000-0002-7075-9384

Germán VARAS ESPINOZA<sup>2</sup>  0000-0001-5426-9839

Valeria GONZÁLEZ HERRERA<sup>1</sup>  0000-0003-2257-1907

### Abstract

The value of scientific knowledge is highly dependent on the quality of the process used to produce it, namely, the quality of the peer-review process. This process is a pivotal part of science as it works both to legitimize and improve the work of the scientific community. In this context, the present study investigated the relationship between review time, length, and feedback quality of review reports in the peer-review process of research articles. For this purpose, the review time of 313 referee reports from three Chilean international journals were recorded. Feedback quality was determined estimating the rate of direct requests by the total number of comments in each report. Number of words was used to describe the average length in the sample. Results showed that average time and length have little variation across review reports, irrespective of their quality. Low quality reports tended to take longer to reach the editor, so neither time nor length were related to feedback quality. This suggests that referees mostly describe, criticize, or praise the content of the article instead of making useful and direct comments to help authors improve their manuscripts.

**Keywords:** Evaluation methods. Feedback (communication). Peer review process. Research articles. Review reports.

### Resumo

*O valor do conhecimento científico é altamente dependente da qualidade do processo com o qual se produz, isto é, do processo de avaliação por pares. Esse processo é uma parte fundamental da ciência que legitima e melhora o trabalho da comunidade científica.*

<sup>1</sup> Universidad de La Serena, Departamento de Artes y Letras. Av. Raúl Bitrán 1305, Colina El Pino, La Serena, Chile. *Correspondencia para/Correspondence to:* O. SABAJ MERUANE. E-mail: <omarsabaj@userena.cl>.

<sup>2</sup> Pontificia Universidad Católica de Chile, Facultad de Letras, Programa de Doctorado en Lingüística. Santiago, Chile.

Support: Fondecyt (1130290) "La interacción socio-discursiva en la construcción colectiva del conocimiento científico: la dinámica interna del proceso de evaluación por pares".

Received on July 14, 2017, final version resubmitted on October 10, 2017, and approved on October 16, 2017.

*Como citar este artigo/How to cite this article*

Cabezas del Fierro, P. et al. Peering into peer review: Good quality reviews of research articles require neither writing too much nor taking too long. *Transinformação*, v. 30, n. 2, p. 209-218, 2018. <http://dx.doi.org/10.1590/2318-08892018000200006>



Nesse contexto, o presente estudo explora a relação entre o tempo de revisão, a extensão e a qualidade da retroalimentação dos relatórios de arbitragem no processo de avaliação, por pares, de artigos científicos. Para esse propósito, foi registrado o tempo de revisão de 313 pareceres de três revistas chilenas internacionais. Para a extensão, foi calculada a quantidade de palavras por parecer, e para a qualidade do feedback foi estimada a quantidade de solicitações diretas (ou de nível 3) pelo total de comentários emitidos em cada parecer. Os resultados demonstraram que o tempo médio e a extensão variavam escassamente nos pareceres, independentemente de sua qualidade. Os pareceres de menor qualidade tenderam a demorar menos em chegar ao editor; portanto, nem o tempo nem a extensão estão relacionadas com a qualidade do feedback. Esses resultados sugerem que os árbitros geralmente descrevem, criticam ou elogiam o conteúdo do artigo em vez de proporcionar comentários úteis e diretos que ajudem os autores a melhorar seus trabalhos.

**Palavras-chave:** Métodos de avaliação. Qualidade de feedback. Processo de revisão por pares. Artigo de pesquisa. Parecer de avaliação.

## Introduction

The Peer-Review Process (PRP) of research articles is a socio-discursive coordinating practice that largely determines the generation and dissemination of scientific knowledge (Sabaj; González; Pina-Stranger, 2016). Given its confidentiality, investigating the PRP is a challenging task that would be almost impossible to study without the contribution of journal editors.

A practical problem concerning the PRP is the fact that authoring, editing, and reviewing are roles fulfilled by the same actors, who, depending on the role they assume, pursue different interests (Squazzoni, 2010). Authors want to publish their papers quickly, but they usually do not respond in a timely manner when they take on the role of refereeing. Referees, on the other hand, may argue that they spend too much time reviewing other people's work without payment or any symbolic recognition, which affects their own productivity. Editors need to publish the best papers in their journals, so they require referees to write clear, high-quality and concise reviews so authors can improve their manuscript version straightforwardly.

Different authors have shown (Bakanic; McPhail; Simon, 1989; Paltridge, 2015; Varas, 2015) that high-quality reviews are uncommon, as they usually contain mixed messages which can induce new authors to misinterpret the intention of reviewers, especially when polite requests are made. In this context, a relevant question is whether feedback quality of review reports is related to their length and to the response time of reviewers. In this study, we aim to investigate these relationships to understand thoroughly the process underlying the dissemination of scientific knowledge.

Time is a critical factor to understand the way scientific knowledge is collectively constructed (Azar, 2004; Graf *et al.*, 2007; Hames, 2007; Sabaj *et al.*, 2015a; Sabaj *et al.*, 2015b). Different international institutions (Committee on Publication Ethics, 2015; Word Association of Medical Editors, 2015) devoted to establishing ethical criteria regarding the PRP recognize time efficiency as an important aspect. In fact, time is relevant for each of the actors participating in the PRP: editors need their journals to be published punctually, and authors want to see their articles published as soon as possible. Both tasks depend on how quickly referees accomplish the task of reviewing.

Several studies regarding the PRP (Gupta *et al.*, 2006; Bornmann; Daniel, 2010; Björk; Solomon, 2013; Lyman, 2013) conceptualize time in a general fashion, distinguishing two main stages of the process, i.e., from Submission to Decision (SD), and from Decision to Publication (DP). A third stage can be obtained by assembling both first and second stages, i.e., from Submission to Publication (SP), which represents the total time of the process. The first stage (SD) includes accepted and rejected articles, while the second stage, ranging from DP, only applies to accepted manuscripts.

Considering the three stages (SD, DP, and SP), Björk and Solomon (2013) found that, for accepted articles, SD average time range tended to be equal to the time of the DP, each representing 50% of total time (SP). Similarly, Gupta *et al.* (2006), describing the editorial final decision time in a medical journal, reported that rejections were consistently faster than acceptances. Bormann and Daniel (2010) showed the same tendency, i.e., acceptances take longer than rejections. As Sabaj *et al.* (2015b) have suggested, this general approach to time in the PRP hides important stages of the process, for example, the selection of the referees' time, notification time, and, most importantly for our present study, the reviewers' time.

Bormann and Daniel (2010) provided specific data regarding the referees' time. These authors showed that the referees' recommendation to publish without alterations was faster (1.93 weeks) than rejections (2.14 weeks) and acceptance with major revisions (2.32 weeks). Analyzing the same stage, Sabaj *et al.* (2015b) showed a similar pattern in two journals on humanities and technology: reviewers' time was longer for rejections than acceptances. On the other hand, in the case of a higher education journal, rejections were faster than acceptances (Sabaj *et al.*, 2015b). Although inconclusive, these data suggest that the editorial final decision and reviewers' recommendation do not follow the same logic concerning time: editors, who are responsible for the final decision, are faster in rejecting articles than accepting them (Gupta *et al.*, 2006; Bormann; Daniel, 2010), but reviewers are faster in recommending acceptance than rejecting (Bormann; Daniel, 2010; Sabaj *et al.*, 2015b). Similarly, Kljaković-Gaspić *et al.* (2003) researched time in the PRP of a small Croatian medical journal, finding that the median review time was 29 days. As proposed in these three studies (Kljaković-Gaspić *et al.*, 2003; Bormann; Daniel, 2010; Sabaj *et al.*, 2015b), in the present research, we conceptualize reviewing time as the period ranging from the moment when the referee receives the article to the time the editor receives the review report.

Other studies have analyzed time more specifically as the period needed by the referee to read and review the article. For example, considering the number of articles reviewed, Yankauer (1990) found that the average review time for the "American Journal of Public Health" was 2.4 hours per paper. Lock and Smith (1990) investigated the time needed for conducting a review by analyzing three samples: a group of pediatricians, a sample of psychiatrists, and a main sample that compiled the two. The authors reported that in the three groups referees spent less than 2 hours assessing a manuscript.

The length of the reviewers' report has not been a specific object of investigation. The data available come from studies in the field of discourse analysis where the length of the evaluation can be found as a descriptive and secondary information. Gosden (2003) analyzed two groups of reviews. The first group corresponding to 22 reports whose publication was conditioned, had an average length of 199 words. The second group included 18 reviews for rejected papers with an average of 185 words. According to these data, there is practically no difference in the association between the length of reports and the type of recommendation. Fortanet (2008) investigated reviews of journals of two disciplines, linguistics and business. In Linguistics, the reports ranged from 180 to 3,214 words with an average of 1,240 words per report. The reviews for the journal of business were, on average, considerably shorter (597 words), ranging from 201 to 1,413 words. Bolívar (2011) analyzed 51 reports of a journal on education. The average number of words per report was 304. Finally, Samraj (2016) showed that reports with major revisions are 21% longer than the rejected ones. Major revision reports had 809 words on average, while the rejected ones had 668 words, contradicting the data provided by Gosden (2003).

Quality measures are always a controversial issue because they depend on assessment instruments, purpose, and audience of that assessment. In the context of the PRP, few studies have considered the quality of reviews. Some of the research conducted on this topic has been in the broad field of medicine. Evans *et al.* (1993) investigated the features of the referees who produced good quality reviews for the Journal of General Internal Medicine. The measure of quality was made using a survey for editors containing four questions to determine: (a) whether the reviewer paid appropriate attention to the importance of the research question, (b) whether he/she commented on key issues, (c) whether he/she commented on the strengths and weaknesses of the research method, and (d)

whether the reviewer made constructive comments on the quality of the writing and the presentation of data (Evans *et al.*, 1993). The results showed that the probability of providing good reviews was higher for younger referees (under 40 years old) who had training in research methods and were affiliated with highly prestigious institutions. Good reviewers were also likely to spend more time conducting the review (more than three hours).

Black *et al.* (1998) also investigated the factors associated with high-quality reviews. The authors concluded that referees trained in epidemiology or statistics were more likely to produce good quality reviews. They also found that there was no association between the editors' quality assessment and the time needed by the referee to return his/her evaluation report to the editor. In addition, Black *et al.* (1998) established that review quality increased along with the time required by referees to write their reports (up to three hours).

Van Rooyen; Black and Godlee (1999) developed and validated an instrument to assess the quality of a review. The instrument was based on the proposal initially made by Evans *et al.* (1993) and Black *et al.* (1998). The final version of the instrument has the following items: importance, originality, method, presentation, constructiveness of comments, substantiation of comments (i.e. the degree to which the referees justified their comments or gave examples to clarify them), the interpretation of the results, and a global item that synthesizes a general quality judgment of the whole revision.

As mentioned above, quality is a difficult issue to deal with. All studies revised (Evans *et al.*, 1993; Black *et al.*, 1998; Van Rooyen; Black; Godlee, 1999) determined the quality of a review as a construct assessed by editors or authors, but none of them define quality as a property of the text itself. In our proposal, we use a discursive construct we have called "feedback quality" of referee reports, thus, defining quality from a textual point of view.

Peer review is a special process since the actors involved interact anonymously. Under these conditions, it is difficult to assert if there is such an entity called "peer". The private nature of the referee report conditions what – and in what terms – can be said. Different studies have shown that comments provided by referees are sometimes useless for authors. Bakanic; McPhail and Simon (1989) showed that due to politeness requirements, reviewers commonly used a positive-negative sequence (Fortanet, 2008; Samraj, 2016) without any direct request to the authors. As Bakanic; Mcphail and Simon (1989) argue, these mixed messages can only induce authors to confusion.

Similarly, Kourilova (1998) and Paltridge (2015) discussed that polite, but ambiguous, and imprecise language used in referee reports can discourage authors, particularly new authors that do not share the same cultural background and mother tongue as the referees. Based on these ideas Stossel, 1985; Bakanic; McPhail; Simon, 1989; Paltridge; (2015), Varas, (2015) developed a discursive model to determine the feedback quality of the review report. Following Stossel (1985), Varas (2015) investigated the relation between the status of the reviewers and his/her discursive behavior. Both studies concluded that status is inversely related to discursive quality.

The developed model (Varas, 2015) is based on two main ideas. Firstly, comments have different levels of direct-indirectness; and secondly, only direct requests contribute to the improvement of the manuscript. Imagine an author reading the following comments:

- 1) Any comments on the equipment used?
- 2) The conclusion might be improved.
- 3) The first two paragraphs of the introduction must be rewritten, including a more specific definition of the term morpheme.

In all the above comments, the author is requested to do something. In (1), the author must assume there is something wrong with the equipment, but he/she does not receive any specific information. In (2), the author can interpret the comment as a command or a suggestion. In (3), the author accurately knows what to do. These

comments correspond to each of the three levels of “feedback quality” (Varas, 2015). Therefore, quality is understood as a function of clarity and directness of the comments used in a referee report. In other words, we understand quality from the point of view of the author, as the level of ease to make the changes needed to improve the manuscript. Comments corresponding to the first and second level were considered poor, obscure, polite and pointless. In contrast, “level three” comments were regarded as high-quality comments since they were less likely to be misunderstood.

As we have endeavored to argue so far, most studies on peer review have related time and feedback quality as secondary variables. In the case of length, investigations have associated the type of recommendation (i.e., acceptance with major revisions or rejections) with the amount of words (Gosden, 2003; Fortanet, 2008; Bolívar, 2011; Samraj, 2016), without paying attention to feedback quality. Something similar occurs with time, as most studies have focused on associating the number of hours or days with the type of recommendation (Azar, 2004; Graf *et al.* 2007; Hames, 2007) without considering the quality of review. In these studies, feedback quality, if considered, is mainly assessed through the perception of the authors and editors (Evans *et al.*, 1993; Black *et al.*, 1998; Van Rooyen; Black; Godlee, 1999). To investigate whether these variables are related, our study endeavors to fill in these gaps by associating time and extension with feedback quality using a model based on discursive patterns. This association is in line with the description of the PRP as a socio-discursive coordinating practice.

## Methodological procedures

A collection of 318 review reports from three well-known international Chilean journals were used as data: *Información Tecnológica* (IT), *Formación Universitaria* (FU), and *Onomázein* (ONO) between 2008-2012. The international character of these journals can be observed as the nationalities of reviewers and authors are not concentrated in local regions. For this research, we only considered review reports on original submissions that editors decided to send directly to referees. All reports were the product of the first round of external revision. These journals were considered for two main reasons: firstly, they cover specific topics and disciplines, such as Engineering and Technology (IT), higher education (FU), and Linguistics (ONO); and secondly, their corresponding editors agreed to participate in our study by offering access to often confidential data (Swales, 1996).

Two of the three journals (FU and IT) use a single-blind peer review system, while the other (ONO) adopts double-blind peer review. The three journals vary in terms of the number of articles and issues published per year. Details about these journals, such as the type of peer review, productivity, and processing time can be found in Sabaj *et al.* (2015b).

The categories analyzed were as follows: (i) Review time, (ii) Length, and (iii) Feedback quality, which we defined as:

(i) Review time: Number of days that the referee took to send the review report back to the editor, which includes the date he/she received both the article and the evaluation form from the editor. This period is the referee’s time of response, which includes reading the article, filling out a form, writing the report, and sending an e-mail back to the editor. Since 5 reports lacked information regarding the review time, the final sample consisted of 313 reports.

(ii) Length: Number of words written by the reviewer, excluding the text of the form itself or fragments taken from the article under evaluation. This number was obtained using Microsoft Word’s word-count tool.

(iii) Feedback quality: Level of directness of comments used in a referee report. This was measured as the rate (percentage) of “level three comments” (Varas, 2015) by the total comments of the report. The analytical procedure used to segment and classify comments into a specific level is described in Varas (2015).

## Results

Table 1 shows the distribution of review reports according to their feedback quality. Each rank represents a quartile of the percentage of level three comments present in a review report. Therefore, reviews including less than 25% of level three comments were considered to have low feedback quality, while reviews including 75% or higher concentration of these comments were considered to have very high-quality feedback.

Low-quality feedback reviews were the most numerous, representing almost two thirds of the total (203 out of 313), with an average of 8.94% level three comments. In contrast, review reports with very high-quality feedback were extremely scarce, representing less than 5.00% of the total reports analyzed (14 out of 313). These reports presented an average of 85.73% of level three comments. High and very high-quality reports represented less than 13.00% of the total. Table 1 shows an association tendency between quantity and quality: quality increases as the frequency of each category decreases.

Table 2 presents the association between feedback quality and length of the reviewers reports in the PRP of research articles. Although the distance between the lowest (406.57) and the highest (631.31) mean values is not considerable, a relationship can be observed in which the reviewers' reports containing comments with high and very high-quality feedback tend to be, on average, shorter than those classified as medium and low quality. Very high-quality feedback also showed the lowest value in the maximum interval (1.566) regarding length. Medium-quality feedback reviews were the wordiest on average (631.31), with a maximum interval (3.965). It was interesting to note that among the lowest feedback quality reports there was an extremely short review consisting of one sentence of seven words.

As shown in Table 3, on average, there was no clear relation between feedback quality and review time, yet a tendency relating high and very high-quality feedback to shorter review times can be observed. Regardless of its quality, on average, it takes a month for a report to be sent back to the editor (32.20). Both high and very high-quality feedback reports are written quickly (1 or 3 days) and never take longer than 87 days. There are no

**Table 1.** Review reports by their feedback quality (% of level three comments).

Feedback quality	n	Minimum	Maximum	Average	Standard deviation
Low	203	0	24.56	8.94	8.18
Medium	70	25	48.84	33.85	6.45
High	26	50	72.00	58.58	7.17
Very High	14	75	100	85.73	8.93
General	313	100	100	22.07	22.26

Source: Prepared by authors (2017).

**Table 2.** Feedback quality and length (number of words) of reviewers' reports.

Feedback quality	n	Minimum	Maximum	Average	Standard deviation
Low	203	7	2,735	573.71	476.45
Medium	70	62	3,965	631.31	599.74
High	26	26	3,269	536.81	637.40
Very High	14	40	1,566	406.57	411.54
General	313	7	3,965	577.09	520.60

Source: Prepared by authors (2017).

**Table 3.** Feedback quality and review time (in days).

Feedback quality	n	Minimum	Maximum	Average	Standard deviation
Low	203	0.00	267.00	33.24	31.41
Medium	70	1.00	87.00	31.74	20.36
High	26	1.00	87.00	24.33	17.85
Very High	14	3.00	69.00	27.16	16.52
General	313	0.00	267.00	32.20	27.53

Source: Prepared by authors (2017).

differences between minimum (1) and maximum (87) values for medium and high-quality feedback reports, yet, on average, both categories show more than 6 days apart. Only low-quality feedback reports were sent to the editor after three months.

## Discussion

As the results have shown, the distance between the lower (406.57) and the higher (631.31) amount of words regarding feedback quality was not considerable. This seems to be due to the high variability of the length of referee reports, which, apparently, would not allow to predict patterns of feedback quality. This instability regarding the extension of referee reports was already evident in the literature. Bolívar (2011), for instance, suggested that the average of words in an education journal was 199, while Fortanet (2008) indicated that, in a linguistics journal, the word average per report was 1.240.

This instability was already identified when correlating word averages with a specific type of recommendation. Gosden (2003), for instance, suggested that, in a journal of hard sciences, the word average in a rejection report was 185, while Samraj (2016) indicated that, in a journal of English for specific purposes, the average was 668. The results obtained in these different studies account for the fact that the report length varies considerably according to the type of discipline.

The journals analyzed published articles in different areas. ONO, for instance, publishes works in Linguistics, Philology and translation. Although most of the reports analyzed belong to the field of Linguistics, the articles ranged between phonetics and phonology, discourse analysis and grammar, disciplines which are quite distant one from the other. The difference among these disciplines may have influenced the fact that the length of reports was not a significant variable when associated with feedback quality.

The results of our study also showed that the relation between time and feedback quality was not significant. Similar to report length, it seems that the review process is also highly variable depending on the discipline, among other factors. Sabaj *et al.* (2015b), for instance, suggested that, in two journals of humanities and technology, rejections took longer than acceptance, while in a journal of education rejections were faster than acceptance. Regarding the number of days for revision, Kljaković-Gaspić *et al.* (2003) showed that, in a medicine journal, the revision average was 29 days, while our results showed that the period of evaluation ranged between 3 and 87 days. These results account for the little stability of the time variable.

These results reveal the difficulty to associate time with the quality of reviews, a problem which had already been acknowledged by some authors. For instance, Evans *et al.* (1993) found that age was associated to good reviews, but not to time. As these authors suggested, "although younger reviewers did spend more time on their

reviews, the multivariable modeling demonstrated that age remained a significant predictor of review quality even after controlling for the time spent on the review" (Evans *et al.*, 1993, p.426).

Black *et al.* (1998) could not find an association between the editors' assessment of review quality and the time taken by reviewers to return their reviews. According to these authors, "there was, in contrast, a clear nonlinear relationship with the time spent by the reviewers on their reviews" (Black *et al.*, 1998, p.233).

As discussed, the results of our study showed that the variation of time and length of the review reports was not associated with feedback quality. The quality of reports seemed to vary according to other factors, for instance, the type of participation that reviewers have in a journal. Varas (2015) identified four types of involvement, i.e., acting as a reviewer once; acting as a reviewer in multiple occasions; acting as an evaluator and author once; and acting as an evaluator and author in multiple occasions. Time and word length seem to vary, but not enough to impact the quality of comments offered by journal reviewers. This might be a further indication that evaluative comments, either bad or good ones (level I, II or III), are more or less stable propositions accounting for the institutionalization of this type of academic discourse. Thus, the fact that the editor receives a poor-quality report, i.e., including obscure, polite and pointless comments, is part of the academic genre. This suggests that the act of writing a review has become or evolved into an 'empty ritual': referees mostly describe, criticize or praise the content of the article instead of making useful and direct comments to help authors improve their manuscripts.

## Conclusion

In this study, the feedback quality of the reviewers reports of research articles was described considering the length of the reports and the review time. From our results, it can be concluded that there is no clear relation between feedback quality, length of reports, and review time, yet some tendencies can be identified: length and time showed to be inversely associated with feedback quality, i.e., shorter and faster reviews tended to be better than longer and slower ones.

These results weaken reviewers concern on the fact that 'reviewing others work is a way of losing productivity'. In fact, conducting a good quality review requires neither writing too much nor taking too long. If we complement these data with other results which have shown that a referee can write a review in about two hours (Lock; Smith, 1990; Yankauer, 1990), the argument against the referees' productivity concern becomes stronger.

The results of this research might be useful to authors, referees, and editors. Editors could use these data to design review guidelines that ensure the feedback quality of review reports. Sabaj *et al.* (2015a) made a proposal in this line that clearly advises reviewers to provide directive, clear, unambiguous comments to authors, and to avoid unnecessary information, vague prose, or descriptions.

Improved guidelines would also help referees have a clearer picture of his/her job. As a virtuous circle, if authors received better feedback, their articles would be certainly improved. Consequently, this would ultimately lead to better publications, which is valuable to the journal editor.

The results could also serve as a warning sign for editors. As our data suggest, when a review report takes longer than 3 months to reach the editor, its feedback quality is always low. For editors, this can mean investing a lot of time and resources and getting a low payback. Thus, a good policy would be that editors ask reviewers to send the report back within a maximum of two months or, otherwise, exclude any report that takes more than 90 days to be returned. As Sabaj *et al.* (2015b) have suggested, having better deadlines could improve the entire process.

Furthermore, an ideal review report would be one containing mainly 'level three' comments and taking less than 3 months to be get back to the editor, as feedback quality dramatically decreases after this period. Clearer

review protocols could reduce length and improve feedback quality, and new editorial policies could set an acceptable period to return the evaluation report to the editor, thus improving the overall quality of the process.

The limitation of the study is that conclusions are difficult to generalize, since data only came from three journals from one country. Therefore, to strengthen the conclusions of research, it would be important to explore the effect of other variables such as discipline, country, prestige of the journal, and level of experience of reviewers.

## Contributors

All authors contributed to the conception, design of the study, data analysis, and final editing.

## References

- Azar, O. Rejections and the importance of first response times. *International Journal of Social Economics*, v. 31, n. 3, p. 259-274, 2004. doi 10.1108/03068290410518247
- Bakanic, V.; McPhail C.; Simon R. Mixed Messages: Referees' comments on the manuscripts they review. *The Sociological Quarterly*, v. 30, n. 4, p. 639-654, 1989.
- Björk, B.; Solomon, D. The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, v. 7, n. 4, p. 914-923, 2013. doi 10.1016/j.joi.2013.09.001
- Black, N. *et al.* What makes a good reviewer and a good review for a general medical journal? *Journal of the American Medical Association*, v. 280, n. 3, p. 231-233, 1998.
- Bolívar, A. Funciones discursivas de la evaluación negativa en informes de arbitraje de artículos de investigación en educación. *Núcleo*, v. 23, n. 28, p. 59-89, 2011.
- Bornmann, L.; Daniel, H. D. How long is the peer review process for journal manuscripts? A case study on Angewandte Chemie International Edition. *Chimia: International Journal for Chemistry*, v. 64, n. 1, p. 72-77, 2010. doi:10.2533/chimia.2010.72
- Committee on Publication Ethics. *Codes of conduct and best practice guidelines*. Eastleigh: COPE, 2015. Available from: <<http://publicationethics.org/resources/code-conduct>>. Cited: Mar. 13, 2016.
- Evans, A. *et al.* The Characteristics of peer reviewers who produce good-quality reviews. *Journal of General Internal Medicine*, v. 8, n. 8, p. 422-428, 1993.
- Fortanet, I. Evaluative language in peer review referee reports. *Journal of English for Academic Purposes*, v. 7, n.1, p. 27-37, 2008.
- Gosden, H. Why not give us the full story? Functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes*, v. 2, n. 2, p. 87-101, 2003.
- Graf, C. *et al.* Best practice guidelines on publication ethics: A publisher's perspective. *International Journal of Clinical Practice*, v. 61, n. 152, p. s1-s26, 2007. doi:10.1111/j.1742-1241.2006.01230.x
- Gupta, P. *et al.* What is submitted and what gets accepted in Indian Pediatrics: Analysis of submissions, review process, decision making, and criteria for rejection. *Indian Pediatrics*, v. 43, p. 479-489, 2006.
- Hames I. *Peer review and manuscript management in scientific journals: Guidelines for good practice*. Oxford: Blackwell Publishing, 2007. doi:10.1002/9780470750803
- Kljaković-Gaspić, M. *et al.* Peer review time: how late is late in a small medical journal? *Archives of Medical Research*, v. 34, n. 5, p. 439-443, 2003.
- Kourilova, M. Communicative characteristics of reviews of scientific papers written by non-native users of english. *Endocrine Regulations*, v. 32, p. 107-114, 1998.
- Lock, S.; Smith, J. What do peer reviewers do? *Jama*, v. 263, n. 10, p. 1341-1343, 1990.
- Lyman, R. L. A three-decade history of the duration of peer review. *Journal of Scholarly Publishing*, v. 44, n. 3, p. 211-220, 2013. doi:10.3138/jsp.44.3.001
- Paltridge, B. Referees' comments on submissions to peer-reviewed journals: When is a suggestion not a suggestion? *Studies in Higher Education*, v. 40, n. 1, p. 106-122, 2015.
- Sabaj, O. *et al.* A new form for the evaluation of scientific articles under peer review. *Revista Argos*, v. 32, n. 62, p. 119-129, 2015a.
- Sabaj, O. *et al.* Relationship between the duration of peer review, publication decision, and agreement among reviewers in three Chilean journals. *European Science Editing*, v. 41, n. 4, p. 87-90, 2015b.
- Sabaj, O.; González, C.; Pina-Stranger, A. What we still don't know about peer review. *Journal of Scholarly Publishing*, v. 47, n. 2, p. 180-212, 2016.
- Samraj, B. Discourse structure and variation in manuscript reviews: Implications for genre categorization. *English for Specific Purposes*, v. 42, p. 76-88, 2016.
- Squazzoni, F. Peering into peer review. *Sociologica*, v. 3, p. 112-115, 2010.
- Stossel, T. Reviewer status and review quality: Experience of the Journal of Clinical Investigation. *The New England Journal of Medicine*, v. 312, n. 10, p. 658-659, 1985.
- Swales, J. Occluded genres in the academy: The case of the submission letter. In: Ventola, E.; Mauranen, A. (Ed.). *Academic writing: Intercultural and textual issues*. Amsterdam: John Benjamins, 1996.

Van Rooyen, S.; Black, N.; Godlee, F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *Journal of Clinical Epidemiology*, v. 52, n. 7, p. 625-629, 1999.

Varas, G. El informe de arbitraje en el proceso de revisión por pares de artículos de investigación: Niveles de retroalimentación según el tipo de evaluador. 2015. 68 f. Tesis (Tesis de Magíster en Estudios Latinoamericanos con mención en Lingüística) – Universidad de La Serena, La Serena, Chile, 2015. Disponible en: <<https://omarsabaj.files.wordpress.com/2017/01/varas-20151.pdf>>.

Acceso: 13 mar. 2016.

Word Association of Medical Editors. *Best practices for peer reviewer selection and contact to prevent peer review manipulation by authors*. New Delhi: WAME, 2015. Available from: <<http://www.wame.org/about/policy-statements#Best%20Practices%20for%20Peer%20Reviewer%20Selection>>. Cited: Mar. 13, 2016.

Yankauer, A. Who are the peer reviewers and how much do they review? *Jama*, v. 263, n. 10, p.1338-1340, 1990.