Reviewing Reviews: An Evaluation of Peer Reviews of Journal Article Submissions

Laura J. Falkenberg and Patricia A. Soranno

Abstract

Publication is the key means by which science is disseminated, with evaluation by journal editors and peer reviewers an important component of the scientific process. Peer reviews are, however, a typically occluded genre of documents not publicly available. Consequently, relatively little is known about peer reviews, including what makes them relevant to editors who are assessing submitted manuscripts. Here we aim to address this knowledge gap by answering the questions: (1) Does reviewer and editor (dis)agreement on the manuscript decision (i.e., accept, minor revision, major revision, reject) relate to how editors assess overall review quality? and, (2) What are the characteristics of review text that distinguish high quality reviews from lower quality reviews for editors? We analyzed 49 reviews of 26 manuscript submissions to *Limnology and Oceanography: Letters*. We found editor perception of review quality was based on review content rather than if there was agreement on the manuscript decision. Specifically, reviews judged by editors to be 'highly relevant' rather than 'sufficient' were typically: longer; included more comments related to study goals, analyses conducted, and resulting claims; and contained more descriptive terms related to the manuscript's importance. Reviewers who consider these factors may produce reviews most relevant to editor decisions.

Introduction

A key feature of contemporary academic publishing is the process of peer review, whereby experts in a discipline become quality control agents over manuscripts submitted to journals in their field. Under this model journal editors elicit reviews from expert peers to inform their decision on whether a manuscript should be accepted for publication or not, and to identify ways in which the submitted manuscript could be improved (Provenzale and Stanley 2006; Mungra and Webber 2010). It is assumed that editors will be better positioned to ensure the quality of their decisions when provided with reviews of high quality.

Given their influence on the decisions made by journal editors, peer reviews clearly form an important genre of academic writing. Despite this importance, peer reviews remain relatively poorly understood compared to other types of texts. This lack of understanding has manifested, in part, because peer reviews are a largely occluded or closed genre of writing not public in nature (Swales 1996). Although there are a range of guidelines available to peer reviewers on the qualities of a good review (e.g., Provenzale and Stanley 2006; Curzon and Cleaton-Jones 2011), these recommendations are typically based on an individual's experiences and preferences, rather than any quantitative analysis. Peer review can be, consequently, a difficult genre of writing to produce, particularly for early career researchers who have little experience with this kinds of writing.

In recent years there has been increasing interest in studying peer review texts to lend insight into the peer-review process. For example, a number of studies have analyzed the parts of submitted manuscripts most often commented on (Bordage 2001; Gosden 2003; Mungra and Webber 2010; Coniam 2012), or the language used in peer reviews (Hewings 2004; Paltridge 2015). In terms of peer review content, analyses have identified

that features often commented on include (among others) those related to technical detail, specifically the association between the claims made and supporting data or the methodology used (Bordage 2001; Gosden 2003; Mungra and Webber 2010; Coniam 2012). Where language has been considered, peer reviews have been found to use more positive than negative adjectives (62% v 38 % respectively; Hewings 2004), with comments often phrased as indirect requests that should be read as directions (Paltridge 2015).

A matter yet to be addressed is how peer reviews are perceived by editors. It is possible that the commonly-identified features of peer reviews may not be those editors find most relevant in making their decisions. In this study, we asked two questions about peer review and editor perception:

1. Does reviewer and editor (dis)agreement on the manuscript decision (i.e., accept, minor revision, major revision, reject) relate to how editors assess overall review quality?
2. What are the characteristics of review text that distinguish high quality reviews from lower quality reviews for editors?

Our goal in this analysis was to identify features of reviews of value to editors, such that we could develop evidence-based guidelines to enhance the quality of peer reviews.

Approach to evaluating reviews

We analyzed reviews submitted to the journal *Limnology and Oceanography: Letters* (*L&O Letters*), which publishes manuscripts that present results, discoveries, or conceptual developments in any area of limnology and oceanography or its integration. *L&O Letters* publishes three types of articles—letters, current evidence, and essays. In this review we focused on the letters submissions, which are short-form articles that present original innovative research advancing knowledge in an area of aquatic science and are the most representative of a typical peer-reviewed article.

At *L&O Letters*, one to three reviewers evaluate all manuscripts, with the majority of manuscripts considered by two reviewers. Reviewers are asked to provide a recommendation from one of four decision categories: accept (the paper is suitable for publication in its current form), minor revision (the

paper will be ready for publication after light revisions), major revision (the paper needs substantial changes such as: expanded data analysis, expansion of literature review, rewriting of sections; however, the analysis, interpretations and conclusions are generally solid and unlikely to change substantially), or reject (the paper does not meet the criteria for publication in *L&O Letters*). In addition, reviewers are asked to include the following components in their review: a summary of the topic of the paper and major conclusions; an overall assessment of the manuscript significance, breadth, and fit; a synthesis of the most critical aspects of the manuscript that should influence the manuscript decision; suggestions to authors to improve the manuscript; a list of any additional comments. The peer reviews, incorporating both the recommended decision and written comments, inform the Associate Editor's decision and comments. All of these are then provided to the Editor-in-Chief. For every review received, a quality assessment can be completed in which reviews are ranked on a scale from 1 to 3: 1, the review is below average; 2, the review is sufficient; 3, the review is highly relevant. Assessments of the reviews are typically completed by the Associate Editors, and occasionally by the Editor-in-Chief if otherwise missing. Thus, the assessments reflect the views of a wide range of editors.

We analyzed 49 reviews of 26 manuscripts submitted to *L&O Letters* between January 2016 and June 2017 for which editors provided a rating of the review relevance. All but two reviews were ranked in the categories of sufficient or highly relevant, with the below average reviews not included in our analysis (due to sample size). We compiled an electronic corpus of all peer reviews meeting the selection criteria. To address our first question about alignment between the decisions of editors and reviewers, we collated: a) the reviewer recommendation for the manuscript, b) the editor decision on the manuscript, and c) the editor-assigned quality assessment of the review. To address our second question, we used a text mining approach in which we identified evaluated ENTITIES and evaluative ADJECTIVES that belonged to the broad quality groupings defined in Hewings (2004) (summarized in Table 1). ENTITIES are components of the manuscript discussed by the reviewer. For example, ENTITIES can be the *manuscript* overall, a particular *point*

made, or a *method* used. Evaluative ADJECTIVES are terms used by the reviewer to describe each ENTITY. For example, evaluative ADJECTIVES could be modifying words such as an *important* manuscript, a *fascinating* point, or a *novel* method. Although we followed an established protocol in evaluating the peer-review text, this approach is just one possible way to deconstruct reviews and interpret the data. We quantified the differences between the sufficient and highly relevant reviews using χ^2 or Student's t-tests.

The (un)importance of recommendation selection

One outcome of the peer-review process is the recommendation made by both the editor and the reviewer: accept, minor revision, major revision, or reject. We found that editor perception of review quality was not related to whether they agreed with the recommended decision of the reviewer (Table 2, $\chi^2_{(1,46)} = 0.26$, $p = 0.61$). This result could occur for a variety of reasons. For example, the judgement of the editor and reviewer regarding the extent of changes that require minor versus major revision may differ, but the key issues identified align. Therefore, our results indicate there is some characteristic of the review text itself that leads editors to distinguish between reviews of contrasting quality.

The importance of the contents of peer reviews

We found the features of peer review text related to editor-perceived quality included the overall length, and the number and type of evaluated ENTITIES and evaluative ADJECTIVES. Specifically, reviews rated as highly relevant were typically longer, and contained significantly more ENTITIES and ADJECTIVES than sufficient reviews (Table 3). These results suggest that editors find reviews most helpful when they are longer and where ideas are elaborated upon, examples provided, and descriptive language used. We do note, however, that there are exceptions to this rule, with some highly relevant reviews being short and containing relatively few ENTITIES and ADJECTIVES. This general result contrasts with common guidelines for scientific writing in which conciseness and avoidance of appraising ADJECTIVES are often promoted (e.g., Grif-

TABLE 1. Classification of evaluated ENTITIES and evaluative ADJECTIVES included in the analysis (derived from Hewings 2004).

Evaluated Entities	
Entity class	**Term**
Paper	Paper, article, manuscript, contribution, research quality, length, level, content, attempt
Expression	Specific wording, section, part, expression, wording, phrasing, choice of words, presentation, style (writing), sentence, paragraph, punctuation, use of language, pronouns, metadiscourse, verb tense, title, editing, language (English), abstract
Claim	Claim, generalization, statement, assertion, discussion, point, argument, case, position, view, suggestion, recommendation, comment, commentary, observation, conclusion, implication, explanation, interpretation, speculation, assumption, reason, justification
Analysis	Finding, result, account, report, description, analysis, term, concept idea, consideration, distinction, classification, categorization, relationship, connection, table, treatment, definition, figure, diagram, chart, comparison, insight, list, appendix, details
Goal	Research question, topic, subject, approach, focus, emphasis, perspective, aim, goal, purpose, issue
Evidence	Data, example, exemplification, sample, evidence, corpus, information, material
Literature	Bibliography, references, literature review, literature survey, overview of the literature
Procedure	Procedure, method, methodology, rationale, theory
Knowledge	Knowledge, awareness, command, familiarity, grasp, understanding

Evaluative Adjectives	
Adjective class	**Term**
Interest	Interesting, original, innovative, fascinating, intriguing, unusual, stimulating, ambitious, attractive, illuminating, rare, novel, pertinent, thought-provoking, unexpected, tedious, uninteresting, unsurprising, conservative, old
Suitability	Good, appropriate, relevant, suitable, excellent, effective, reasonable, satisfactory, strongest, successful, odd, weak, stronger, inappropriate, unfortunate, poor, unsuitable, irrelevant, inadequate
Comprehensibility	Clear, succinct, easy, straightforward, concise, confusing, clearer, unclear, awkward, abstract, distracting, disjointed, difficult, disconnected
Accuracy	True, accurate, careful, consistent, solid, valid, cogent, plausible, principled, systematic, wrong, anecdotal, inconsistent, categorical, contentious, contradictory, inaccurate, unwarranted, bumpy, disingenuous, extreme, imprecise, loaded, relaxed, speculative, unfocussed
Importance	Useful, valuable, important, helpful, substantial, practical, worthwhile, salient, crucial, meaningful, prominent, meaningless, unhelpful
Sufficiency	Sufficient, thorough, comprehensive, small, redundant, thin, briefer, fuller, narrow, underdeveloped, excessive, simplistic, brief, large, oversimplified, repetitive, lengthy, minimal, narrower, occasional, repetitious, superficial, unexplained, unsupported
Praiseworthiness	Impressive, admirable, laudable, disappointed, puzzled, disappointing, uncertain, frustrating, unsure
Perceptiveness	Sophisticated, insightful, intelligent, sensible, informed, perceptive, unaware

fies et al. 2013). Peer-review writing may, therefore, require authors to adopt different strategies and styles than are commonly recommended for academic writing.

We found that higher quality reviews had more ENTITIES belonging to the classes of: ANALYSIS, CLAIM, and GOAL (Fig. 1). The finding that relevant reviews contain more comments in the ENTITY class of ANALYSIS indicates editors value consideration of the details which underlie the conclusions drawn (e.g., how data are represented, analyses conducted). Rather than representing an increase in the number of detailed comments (i.e., specific line edits), these terms were typically included in higher-level comments (i.e., providing evidence to statements of key interest). Further, the frequent use of ENTITIES belonging to both the CLAIM and GOAL classes shows that highly-relevant reviews emphasize broader issues of how the manuscript fits within the wider scientific context.

TABLE 2. The proportion of reviews that are classed as highly relevant or sufficient, and those for which the reviewer recommendation and editor decision are the same/different.

		Relevance of review		
		Highly relevant	**Sufficient**	**Total**
Agreement between reviewer recommendation and editor decision	Yes	41 %	35 %	76 %
	No	11 %	13 %	24 %
	Total	52 %	48 %	

TABLE 3. Characteristics of the reviewed reviews.

Review characteristic	Highly relevant				Sufficient				Comparison		
	Mean	SE	Max	Min	Mean	SE	Max	Min	df	T	p
Words	963	99	2877	291	453	48	645	194	45	2.94	**0.01**
Evaluated Entities	37	4	105	2	20	3	38	5	45	2.23	**0.03**
Evaluative Adjectives	12	1	32	0	6	1	10	0	45	2.49	**0.02**

Those characteristics in **bold** were significantly different between highly relevant and sufficient reviews as indicated by Student's *t*-tests
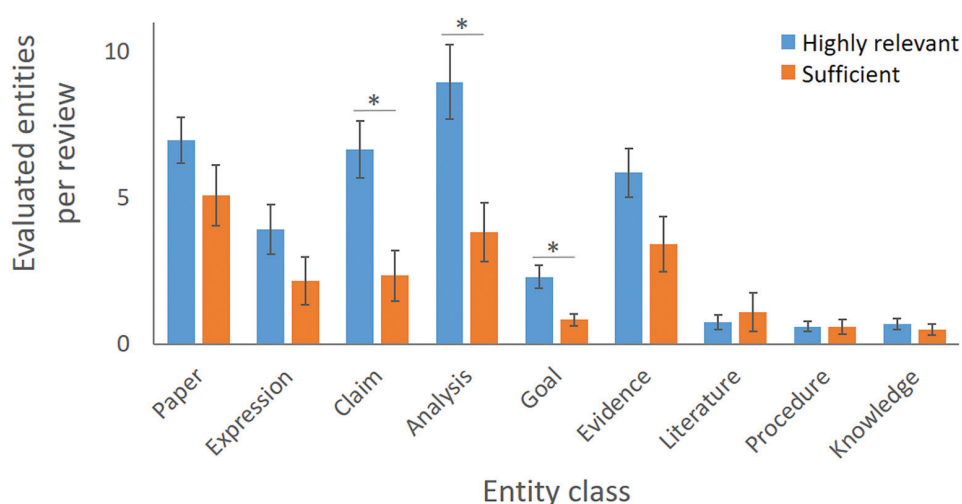
FIG. 1. Evaluated ENTITIES. *indicates significant differences in that ENTITY class identified with Student's *t*-tests.
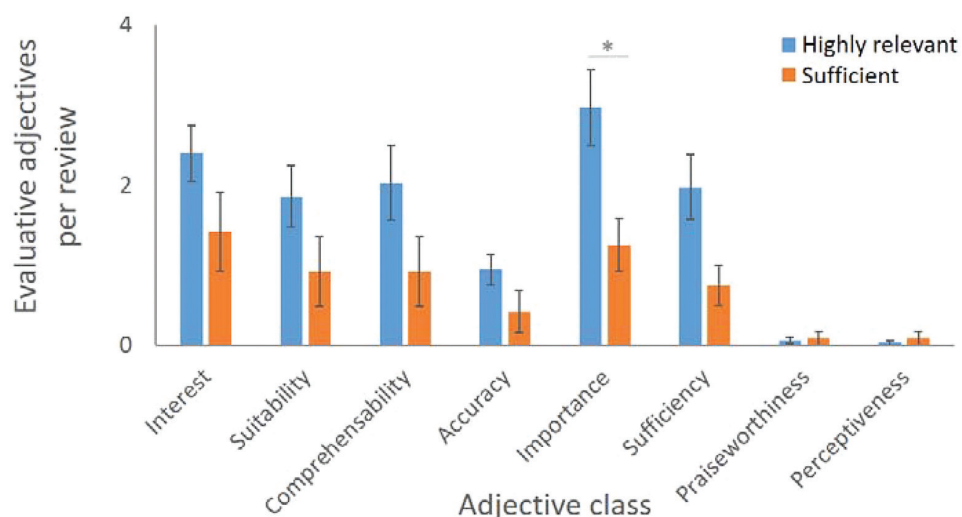
FIG. 2. Evaluative ADJECTIVES *indicates significant differences in that ADJECTIVE class identified with Student's *t*-tests.

Terms from these classes were typically used in comments on how the manuscript was stepping beyond existing literature, and the potential contribution of presented results. It seems logical that such components of a review would be valuable to an editor making a decision regarding a manuscript's scientific integrity, significance, and contribution to the field. While evaluation of these components of a review typically require terms from the ENTITIES discussed above, other terms may also be relevant for certain manuscripts. For example, if a manuscript cited no LITERATURE to provide broader context, this would be a key ENTITY to emphasize despite it being little represented in highly relevant reviews overall. Consequently, we suggest authors of peer reviews consider these important ENTITY terms for evaluating both the detail and the wider context of manuscripts.

We found that highly relevant reviews also had more evaluative ADJECTIVES compared to sufficient reviews, indicating that use of terms from all ADJECTIVES classes are valued by editors. Although this pattern was found across all ADJECTIVES classes, only the class of IMPORTANCE was significantly different between the highly relevant and sufficient reviews (Fig. 2). The emphasis on the IMPORTANCE class indicates editors value comments regarding the potential contribution of a submitted manuscript to the discipline. This result is supported by the higher numbers of the ENTITY class of CLAIM in highly relevant reviews described above (which considers another aspect of the potential contribution of the presented results). We suggest this finding re-emphasizes our recommendation that peer reviewers consider not only the detail of the manuscript they are considering, but also the wider context in which it would exist if published.

How can peer reviews be most relevant to editors?

Overall, our results suggest that editors do not base their perception of review quality on a reviewer's recommendation, but rather the content of the review itself. Highly-relevant reviews were typically longer and contained more evaluated ENTITIES and evaluative ADJECTIVES. We suggest that these kinds of results can inform peer-review writing guidelines (e.g., Box 1). We propose that these strategies could be broadly applicable to peer reviewing in other journals. Although our conclusions are drawn from an analysis of a single journal, reviews and editor assessments reflected the views of a range of academics who are involved in peer review for many journals. Together these strategies will prompt peer reviewers to consider manuscripts more complexly. While it can be easy to focus on certain evaluated ENTITIES and evaluative ADJECTIVES, editors

use a diverse range of terms to inform their decisions and feedback to authors, with this intricacy often reflected in highly relevant peer reviews.

Acknowledgments

BOX 1. Tips for producing a manuscript review relevant to journal editors

1. *Remember your recommended decision is just part of the review*—while this should be selected carefully, the recommendation is not the main factor that determines the relevance of a review
2. *Be expansive in your review such that you evaluate all relevant aspects and provide examples where relevant*—although much scientific writing is about being concise, highly relevant reviews are often longer than their sufficient counterparts
3. *Evaluate* ENTITIES *that: (a) reflect the detail of the manuscript*—highly relevant reviews address the specifics from which everything else is drawn, with these often covered by ANALYSIS terms—*and (b) indicate the wider scientific context*—the broader landscape can be reflected by ensuring terms from the CLAIM and GOAL classes are included
4. *Be descriptive*—generous use of evaluative and informative ADJECTIVES can help an editor understand why you are commenting on a particular feature of a manuscript; specifically, be sure to highlight where the IMPORTANCE of a manuscript lies
5. *Use an approach relevant to the manuscript under review*—the traits that characterize a highly relevant review will vary, and reviewers will benefit from considering which ENTITIES or ADJECTIVES (e.g., Table 1) should be emphasized for each specific manuscript

---

References

Bordage, G. 2001. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. Academic Medicine **76**:889–896.

Coniam, D. 2012. Exploring reviewer reactions to manuscripts submitted to academic journals. System **40**:544–553.

Curzon, M., and P. Cleaton-Jones. 2011. Reviewing scientific manuscripts. European Archives of Paediatric Dentistry **12**:184–188.

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**:87–101.

Griffies, S. M., W. A. Perrie, G. Hull. 2013. Elements of style for writing scientific journal articles. Publishing Connect, Elsevier.

Hewings, M. 2004. An 'important contribution' or 'tiresome reading'? A study of evaluation in peer reviews of journal article submissions. Journal of Applied Linguistics **1**:247–274.

Mungra, P., P. Webber. 2010. Peer review process in medical research publications: Language and content comments. English for Specific Purposes **29**:43–53.

Paltridge, B. 2015. Referees' comments on submissions to peer-reviewed journals: When is a suggestion not a suggestion? Studies in Higher Education **40**:106–122.

Provenzale, J. M., R. J. Stanley. 2006. A systematic guide to reviewing a manuscript. Journal of Nuclear Medicine Technology **34**:92–99.

Swales, J. 1996. Occluded genres in the academy. In: Ventola E, Mauranen A (eds) Academic Writing: Intercultural and Textual Issues. John Benjamins, Amsterdam

Laura J. Falkenberg, Norwegian Institute for Water Research, Thormøhlens gate 53 D, 5006 Bergen, Norway; laura.falkenberg@outlook.com

Patricia A. Soranno, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI