Information and Software Technology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

A community's perspective on the status and future of peer review in software engineering

Lutz Prechelt[*,a], Daniel Graziotin[b], Daniel Méndez Fernández[c]

[a] Freie Universität Berlin, Berlin, Germany
[b] University of Stuttgart, Stuttgart, Germany
[c] Technical University of Munich, Garching, Germany

ARTICLE INFO

Keywords:
Survey
Peer review
Open review
Double-blind review
Reward

ABSTRACT

*Context:*Pre-publication peer review of scientific articles is considered a key element of the research process in software engineering, yet it is often perceived as not to work fully well.

*Objective:*We aim at understanding the perceptions of and attitudes towards peer review of authors and reviewers at one of software engineering's most prestigious venues, the International Conference on Software Engineering (ICSE).

*Method:*We invited 932 ICSE 2014/15/16 authors and reviewers to participate in a survey with 10 closed and 9 open questions.

*Results:*We present a multitude of results, such as: Respondents perceive only one third of all reviews to be good, yet one third as useless or misleading; they propose double-blind or zero-blind reviewing regimes for improvement; they would like to see showable proofs of (good) reviewing work be introduced; attitude change trends are weak.

*Conclusion:*The perception of the current state of software engineering peer review is fairly negative. Also, we found hardly any trend that suggests reviewing will improve by itself over time; the community will have to make explicit efforts. Fortunately, our (mostly senior) respondents appear more open for trying different peer reviewing regimes than we had expected.

1. Introduction

For our purposes, peer review is the practice by which a publication venue sends an article to several expert colleagues (the peers) for review before it is accepted for publication (or not). Although a few venues recently started trying out a different approach (e.g., [9,22]), this basic model of *pre-publication peer review* is usually considered a cornerstone of quality assurance in the scientific process, in software engineering and beyond [14].

This article attempts to understand what is currently working well or not-so-well about peer review in software engineering (SE) and how this might change in the next 20 years.

1.1. Variants of peer review

The acceptance decision may be made after just one round of reviewing (single-stage peer review[1]), typical for conferences, or after multiple rounds with improvements of the work (multi-stage review[2]), typical for journals.

Usually, the authors do not know the identity of the reviewers (blind review). Reviewers might know the identity of the authors (single-blind review) or not (double-blind review). Only rarely do the authors get to know the names of reviewers (non-blind review, zero-blind review) or does the public get to see the content of the reviews (open review, public review).

1.2. Issues with peer review

Informally, researchers often criticize peer review as not doing its job properly and indeed the practice has various inherent problems, for instance:

- Reviewers will not always be competent to properly review a particular work, and often provide inconsistent reports [3,24].

* Corresponding author.
E-mail addresses: prechelt@inf.fu-berlin.de (L. Prechelt), daniel.graziotin@informatik.uni-stuttgart.de (D. Graziotin), daniel.mendez@tum.de (D.M. Fernández).

[1] authors may be allowed to comment on the reviews: rebuttals.
[2] always with rebuttals.

http://dx.doi.org/10.1016/j.infsof.2017.10.019
Received 23 June 2017; Received in revised form 26 October 2017; Accepted 28 October 2017

- Reviewers will sometimes be biased against certain aspects of the work: methods, technology, goals, etc. [1].
- Reviewers may, protected by their anonymity, abuse their power to inhibit the publication of lines of work that compete with their own [11,24].
- Reviewing can be viewed as contributing little to the reviewer's reputation and so reviewer motivation can be lacking and reviewing be done rather sloppily [28].

Because of issues like these, other fields (most prominently in the biomedical realm) have long worked to understand the status of peer reviewing and how to improve it [15]. For instance, such research has produced strong evidence that double-blind reviewing will lead to results that are less biased than with single-blind reviewing, e.g. Budden et al. [7], a fact that is now also being picked up in software engineering [2]. But beyond that, software engineering venues are not, so far, particularly prone to experimentation with possible improvements to the peer reviewing regime. In light of the above issues, this might be a pity.

For instance, the high-class health journal The BMJ (acceptance rate 7%) not only performs reviews zero-blind (that is, reviewers sign their reviews), they also publish the reviews along with accepted articles (open reviewing, BMJ [4]); there is no comparable software engineering venue doing anything as radical.

1.3. Research questions

Our perspective is understanding and then improving the peer review process. We designed our survey along the following research questions. Results and discussion will be structured mostly into one section per research question.

Section 5: What do authors and reviewers perceive to be the purposes of peer review? Which are more important than which others?

Section 6: How well do they perceive peer review to work today (in the sense of producing valid and helpful reviews) and why?

Section 7: How much should reviewers and authors be blinded?

Section 8: Which aspects of reviewing should be public?

Section 9: Should reviewers be compensated for their work? How?

Section 10: What changes to the current reviewing regime should be performed?

Section 11: How might the answers to each of the above questions change in the next few decades?

1.4. Research contribution

Our article makes two research contributions: First, it characterizes the attitudes of mostly senior members of the ICSE[3] authors-and-reviewers community with respect to the research questions. Second, it predicts how these attitudes will likely be different for a similar sample of people in the future, several decades away.

1.5. Structure of this article

After reviewing related work (Section 2), we will present our method: The survey population (Section 3.1), the survey instrument (Section 3.2), the execution of the survey (Section 3.2), our data analysis techniques (Sections 3.3 and 3.4), and the resulting public data archive (Section 3.5). Then, we discuss the respondent demographics (Section 4) before presenting the results structured according to our list of research questions (Sections 4 – 11). We then discuss our study's limitations (Section 12) before we conclude (Section 13).

[3] International Conference on Software Engineering

2. Related work

We organize this section along the research questions from Section 1.3. What sets our study apart from other survey work in the area is the use of open questions and qualitative analysis. While we reference various related work, we consider two large scale surveys of peer reviewers attitudes across many disciplines as our baseline background material upon which we frame our study primarily: First Mulligan et al. [14] with 4037 respondents, second Ross-Hellauer et al. [20] with 3062. The latter, organized by OpenAIRE, an Open Access collaboration project, is special in that 76% of respondents reported to have participated in open reviewing previously; an unusual population. We found only one reviewing study in the software engineering literature [2], also a survey.

Purpose of peer review: Weller [27, p.xii] proposed a concise characterization: "The valid article is accepted, the messy article cleaned up, and the invalid article rejected". The Mulligan et al. [14] survey found the main perceived purposes to be (in this order): to improve the quality of published papers; to determine their originality; to select the best possible manuscripts for a journal. Our work will ask the question also beyond predefined answer categories and ask for elaboration.

How well does peer review work today: The Mulligan et al. [14] survey had 69% of respondents report high or very high satisfaction. When asked what aspects of their articles were improved the most through peer review, respondents mentioned the introduction most (90%) and statistical methods least. Our work will ask about percentages of good, mediocre, or bad reviews received and about specific positive and negative peer review experiences to provide a more detailed picture.

Blinding: Much discussion has happened lately on how much anonymity should be in the peer review process [8,12]. Empirical research has found interesting effects from double-blind reviewing. For instance, Budden et al. [7] found that more articles of female researchers were accepted after the journal Behavioral Ecology adopted double-blind review (but not in other journals that did not). Laband and Piette [13] found for a sample of economics journals (and controlling for several confounding factors) that articles accepted after single-blind review were cited less often than articles accepted after double-blind review. As for software engineering, Bacchelli and Beller [2] survey how double-blind peer review is perceived by the ICSE community and find that about half of the respondents believe all software engineering venues should switch to double-blind reviewing. Seeber and Bacchelli [23] investigate bibliographic data from 71 of the 80 largest computer science conferences of 2014 and 2015 and find evidence that newcomers (people who have not previously published at that conference) get a smaller share of a conference when single-blind reviewing is used compared to conferences using double-blind reviewing.

The Mulligan et al. [14] survey respondents did not like the prospect that their names be made visible to the authors (8% more likely to be willing to review under such circumstances, 51% less likely) or to the readers (18% and 45%). In the OpenAIRE survey, 67% of respondents believed zero-blind reviewing would make reviewers less inclinced to provide a review and 44% believed it would improve review quality; 65% believed it makes strong criticism less likely [20]. Our study will ask for degrees of agreement with double-blinding and zero-blinding.

Publicness: Support for the review reports to be published alongside the accepted paper was low (11% more likely and 58% less likely) in Mulligan et al. [14]. Similar percentages were found for the possibility of disclosing names to authors only (8% and 51%) and for having the reviewer names only published alongside the article (18% and 45%). Even in the OpenAIRE survey, 52% of respondents expect reviewers to become less inclined to review, although 65% expect published reviews to be useful for readers, 60% expect an increase in review quality, and 45% expect authors to become more inclined to submit to such journals.