EQUAP^2

Evaluating the Quality Assurance Process in Scholarly Publishing

Final Report

February 6th 2023

https://doi.org/10.5281/zenodo.7612114

Konsortium der Schweizer Hochschulbibliotheken

Swiss Library Network for Education and Research

Impressum

EQUAP^2 – Evaluating the Quality Assurance Process in Scholarly Publishing

Author

Dr. Robert Neumann (TU Dresden)

Project Team

Silke Bellanger[a] (UB Basel), ORCID: 0000-0003-1952-1933
Marcel Erfurth[a] (SLUB Dresden)
Josephine Hartwig[a,c] (SLUB Dresden), ORCID: 0000-0002-5336-6853
Anne-Christine Günther[a] (ULB Darmstadt), ORCID: 0000-0002-3152-3846
Ulrike Kändler[a] (TIB), ORCID: 0000-0002-4225-440X
Dr. Robert Neumann[a,b,d] (TU Dresden), ORCID: 0000-0001-6728-4288
Gregor Reiss[a] (University Library TU Munich), ORCID: 0000-0003-2540-6093
Nicole Rosenke[a] (ULB Darmstadt), ORCID: 0000-0001-7597-7885
Dirk Verdicchio[a] (UB Bern), ORCID: 0000-0001-7297-9009
Hagen von Hermanni[a,b] (TU Dresden), ORCID: 0000-0002-8709-9394
Michael Wohlgemuth[a,c] (SLUB Dresden), ORCID: 0000-0002-7280-5284

Contact:

EQUAP2@slub-dresden.de

Funding Statement

The EQUAP^2 survey was financed by the university libraries of TU9 German Universities of Technology, the University/Central Library of Zurich, ZHAW Zurich University of Applied Sciences, University Library of Basel, University Library of Bern, University of Geneva, Consortium of Swiss University Libraries and the Swiss Library Network of Education and Research.

[a]conceptualized the questionnaire; [b]performed the data analysis; [c]coordinated the project; [d]wrote the report

Key Results

- The expectations towards the peer review process are very homogeneous across all disciplines. The most important criteria for quality assurance are (1) transparency with regard to funding sources, (2) transparency with respect to possible conflicts of interest and (3) the professional expertise of the reviewers and the review process in general.

- From the point of view of editors and reviewers, time represents an important quality feature, with very short and very long review times tending to be seen as a sign of poor publication practices. However, short review and response times are highly appreciated by authors.

- The evaluation of the actual perception of the status quo reveals that there is a high degree of variation in the perceived fulfillment of scientific standards by publishers among the scientists surveyed. Particularly with regard to the processing times specified for the publishers, the information provided by editors and reviewers indicates considerable differences.

- There is some evidence for varying entrepreneurial strategies by publishers with regard to the assignment of articles to editors and to the selection of suitable reviewers. Open access (OA) publishers appear to be using algorithm-based procedures to automatically streamline the review process. Concerning this practice, our factorial design about the quality dimensions of journals shows that researchers perceive such procedures as being less reputable.

- Overall, the survey generates the insight that the perception and assurance of quality in peer review processes is subject to various conflicts regarding goals and interests. In contrast to editors and reviewers, authors are less critical of the various practices of publishers.

Contents

1 Introduction

In addition to publication and knowledge dissemination, the selection and peer review of scientific articles are among the central contributions of journals to scholarly communication. The quality assurance processes for journals are also closely linked to the other mechanisms of evaluation and assessment of scientific performance. In the context of institutional funding of publications as well as the provision of information, the criteria of quality assurance by publishers and journals play a central role. This is particularly evident in discussions about publishers, whose business practices, pricing policies and quality assurance processes are being questioned by an increasing number of researchers.

Common bibliometric indicators such as the journal impact factor are not suitable for mapping the diverse dimensions of peer review quality. While prior efforts to assess the quality of the peer review process were motivated by the intrinsic interest of individual journals' editorial boards (see e.g. Weber et al. 2002; Black et al. 1998), we are interested in the concrete experiences of researchers in their roles as authors, reviewers and/or editors across the entire scientific journal landscape and in comparing these experiences with their expectations. Any discrepancies between the experiences and expectations open up the possibilities to derive both publisher-specific statements on quality assurance and evidence-based insights to inform and improve acquisition decisions by scientific libraries.

Therefore, the EQUAP^2 survey pursued three main objectives:

1. Identify scientists' expectations regarding the quality dimensions of the peer review process
2. Investigate whether scientists' experiences match the expectations, depending on the scholarly publisher in charge of the review process.
3. Scrutinize the relevance of different dimensions of the quality assurance process, across disciplines as well as publishers.

First, the EQUAP^2 survey focused on publishers with the highest number of publications, irrespective oft the OA Status oft he journals (closed access, hybrid or fully OA). It also included publishers who are prominent representatives of the group of OA publishers who exclusively publish journal articles in electronic form as OA. The list of publishers at the center of our attention were Elsevier, Springer Nature, Wiley, Informa (Taylor & Francis), MDPI, Frontiers, Sage Publications, PLOS, Oxford University Press (OUP), Copernicus and De Gruyter. Second, the experiences of the scientists with regard to the review process were investigated by distinguishing the subjective experiences from their roles during the review process. Hence, we asked whether they had served as an editor, reviewer or author in the last twelve months and tailored the questions to the specific challenges and requirement within their respective roles.

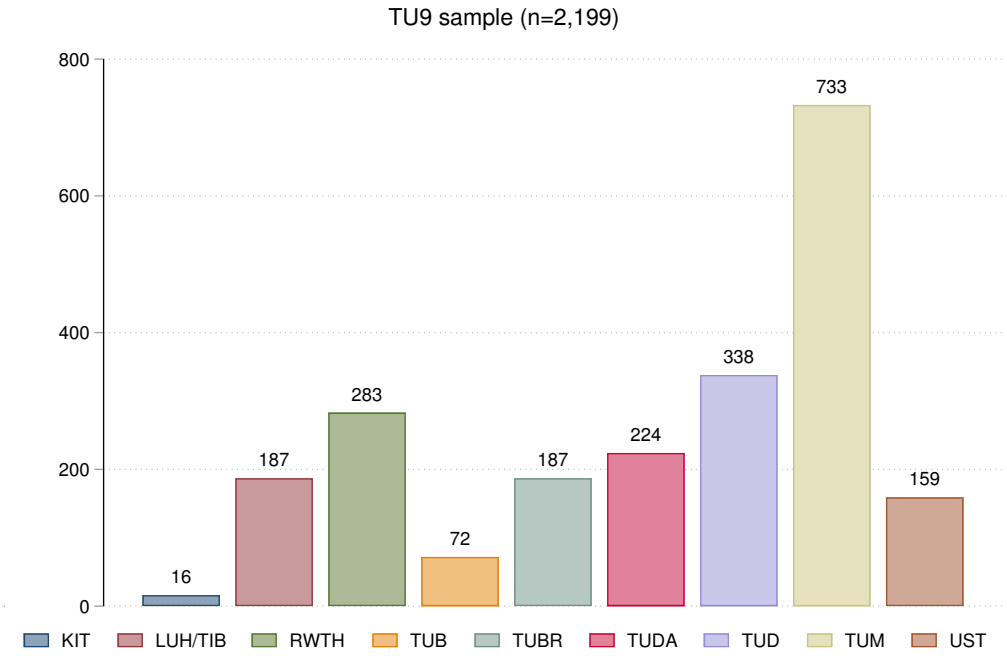


Figure 1: Total number of completed interviews at TU9 universities

2 Study Design

The EQUAP[2] survey took place in Germany and Switzerland between 26 April and 31 August 2022. The interviews were conducted via a bilingual web survey at 26 universities and research institutes. Initially, the EQUAP[2] survey was joint project of the libraries affiliated to the TU9 German Universities of Technology[1] and five institutions from Switzerland.[2] In addition, twelve German and Swiss universities and research institutes joined the project and allowed the dissemination of invitation e-mails to their scientists.

Across all institutions, the target population included all scientists working at the respective university / research institute who could be contacted via an affiliated e-mail address (i.e., emeritus professors or adjunct professors were also allowed to be contacted). Depending on the regulations and data restrictions at the respective institutions, participating institutions either contacted all scientists based on an existing sampling frame of addresses or performed random sampling from a sampling frame. If there was no access to a suitable sampling frame, participating universities were asked to invite as many scientists at their university as possible (the "more-is-better" assumption), without applying any special selection procedure (as in the case of a full sample or a random selection). In this case, mailing lists and newsletters were used to contact scientists at the institutions. Lastly, if no other approach was feasible, a snowball technique was applied whereby the total number of contacts can usually not be determined. Gross and net sample sizes for the different universities are reported in Table A.1 in the Appendix.

[1] TU9 is composed of the Karlsruhe Institute of Technology (KIT), Leibniz University Hannover/Leibniz Information Centre for Science and Technology and University Library (LUH/TIB), RWTH Aachen University (RWTH), TU Berlin (TUB), TU Braunschweig (TUBR), TU Darmstadt (TUDA), TU Dresden (TUD), TU Munich (TUM) and the University Stuttgart (UST).

[2] Institutions from Switzerland were: University/Zentralbibliothek Zurich (UZH), the ZHAW Zurich University of Applied Sciences (ZHAW), the Universitätsbibliothek Basel (UNIBAS), the Universitätsbibliothek Bern (UNIBE) and the University of Geneva (UG).

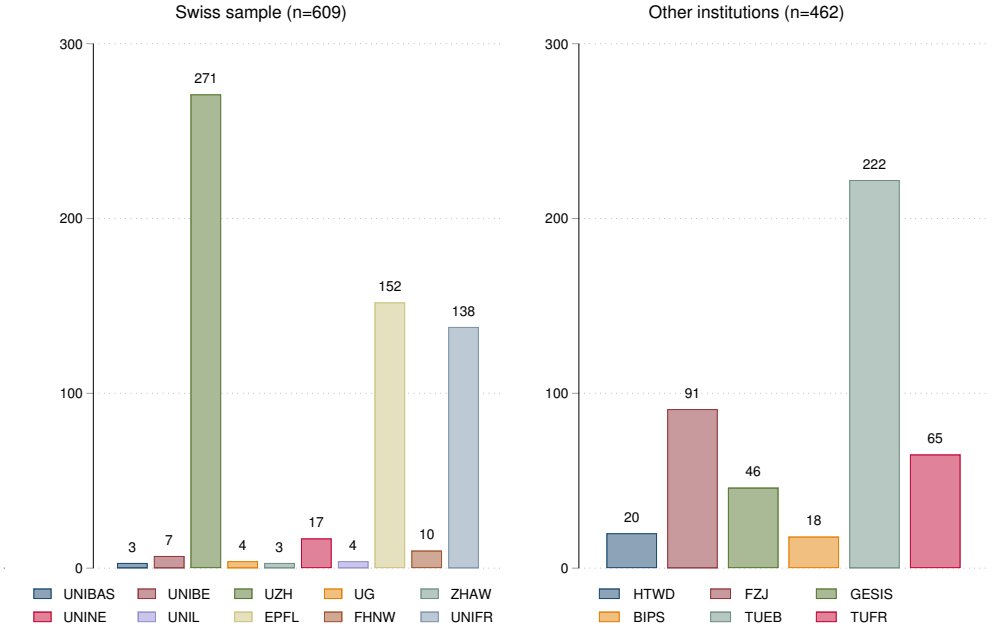


Figure 2: Total number of completed interviews at Swiss universities (left panel) and other institutions (right panel)

3 Sample Composition

All in all, we were able to collect data from more than 3,270 completed interviews across these 25 institutions. Due to the different modes of inviting scholars to participate in the surveys, we are not able to compute response rates in accordance to AAPOR standard guidelines (adapted to the German context, see Stadtmüller et al. 2019). Figure 1 and Figure 2 report the total number of completed interviews across the German TU9 universities as well as Swiss and other institutions. The institutions applied either random sampling from the staff registry, were able to contact their entire scientific staff or had to rely on invitations via newsletters or mailing lists with an unknown number of subscribers. The reasons for these differences were due to the different administrative barriers in contacting researchers via a personalized e-mail invitation. We provide an overview about the modes of inviting researchers at the respective institutions in Table A.1 in the Appendix.

Two aspects of our sample composition are of interest: First, the distribution of scientists across various disciplines. Figure 3 presents the composition of our pooled sample of all institutions with respect to the disciplinary background of the scientists. It is easily observable that the survey was conducted at universities with a curricula focused on Natural Sciences, Engineering and Biology/Pharmacy. The high portion of item non-response is unfortunate; all other disciplines show more or less equal distributions at a rate that allow for quantitative analyses.

Second, in the second part of the questionnaire, we draw a distinction between the respondents according to their roles of serving either as an editor, a reviewer and/or author. This information was collected following the answers of the respondents to the question about their roles within the peer review process in the last twelve months prior to the survey. As our research interest was concerning the experiences of those scientists who were actively involved in the peer review process, we applied a filter after roughly two thirds of the questionnaire where respondents answered in their role as editors in case they served as one in the last year, irrespective if they also served as reviewer or author. If they did not serve as editor, but as reviewers or authors, respondents had to answer the questions concerning

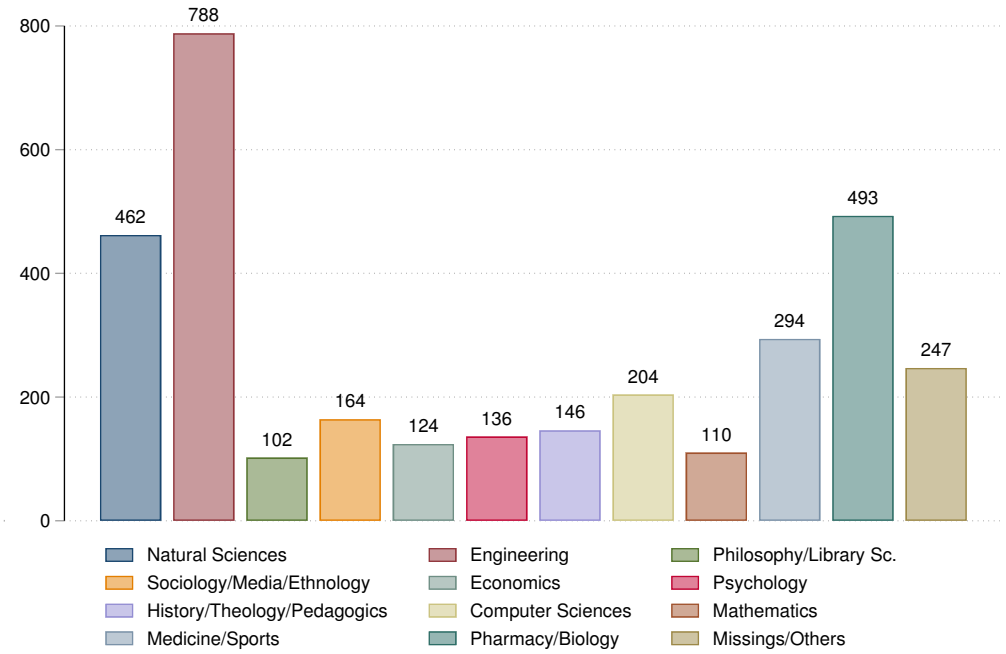


Figure 3: Total number of respondents, differentiated by disciplines (N= 3, 270)

their role as reviewers. If neither the role of editor nor reviewer was selected, respondents filled out the questionnaire that covered the perception as authors.[3] This generated role-specific answers about the quality assurance process from 618 editors, 2,141 reviewers and 450 authors. That is, we were able to connect all of their responses to the specific publisher they worked for. Not surprisingly, the editors have the highest average age (49.3 years) followed by the group of reviewers (42.3 years) and authors (32.1 years).

Finally, we have to deal with different degrees of item non-response throughout the survey. In the first part, non-response to questions about the expectations towards the quality assurance process was negligible. After the filter question about the roles of the respondents, varying sample sizes are the results of respondents either withholding their role, the publisher or the journal they served for. Where possible, we report the sample sizes of the items either individually or report average sample sizes across item batteries.

[3]Respondents who revealed multiple functions were asked at the end of the full questionnaire whether they would be able to also answer voluntarily in their other role they previously indicated.

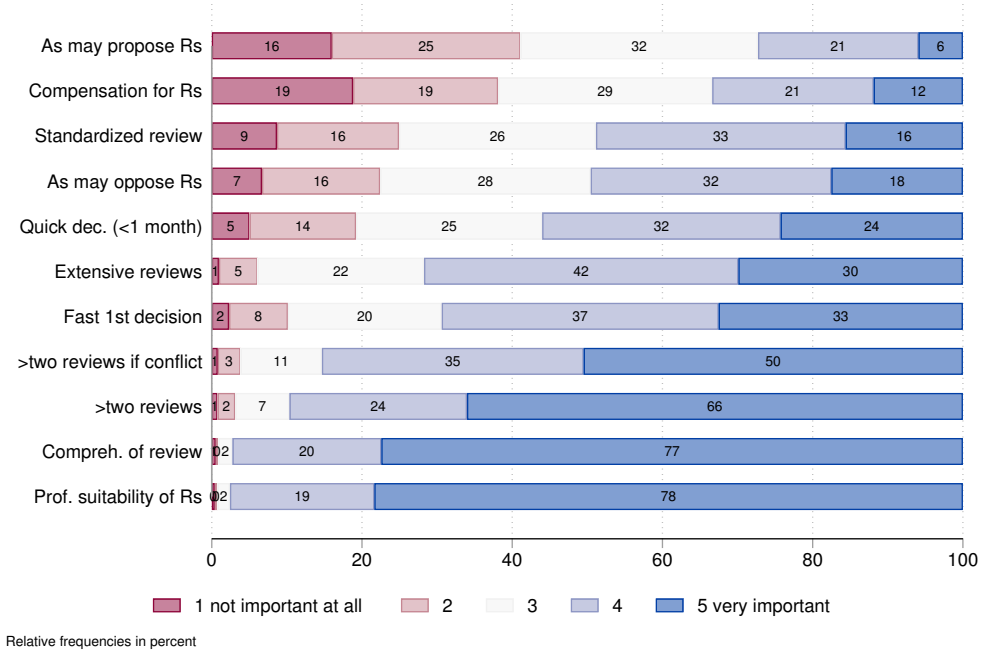


Figure 4: Expectations towards the peer review process to assure best practice (mean number of answers $\bar{n} = 3,261$).

4 Results

The analytical strategy is as follows: First, the main analysis of the expectations towards the review process is performed on a pooled data sample consisting of all interviews. Second, we perform discipline-specific analysis with regard to the expectations of researchers and publisher-based analysis for the perceptions of the status quo. For analytical reasons and to prevent the de-anonymization of individuals, we refrain from institution-specific analyses.[4] Third, because respondents either belong to their respective discipline or manuscripts have been reviewed, rejected or printed within the outlets of specific publishers, our responses can and should be analyzed within the framework of multilevel modeling. The reason is that given that individual answers are nested within disciplines and/or publishers, we have both within-group and between-group variation that both need to be accounted for. This also holds true for our factorial survey design (see Section 4.2), where each respondent gave several answers to two review vignettes and three decision vignettes, hence individual answers to these scenarios are nested within respondents. Finally, the answers about the perception of the status quo are analyzed separately, depending on the role the respondents served in the twelve months prior to the survey.

4.1 Dimensions of the Quality Assurance Process

To assess scientists' expectations about the appropriate way to ensure the quality assurance process, the survey asked the respondents to rate the importance of several dimensions of the review process. Dimensions of interest included items on the organizational process of the entire review process, the scope/length of reviews, the time to arrive at a first and/or final decision and the professional expertise

[4] As will be shown, the expectations of scholars are very homogenous across all the scientific disciplines. That is why it does not necessarily make sense to expect any variation in expectation conditional to institutional affiliation, as these strata only represent smaller sub-samples of the entire disciplines.

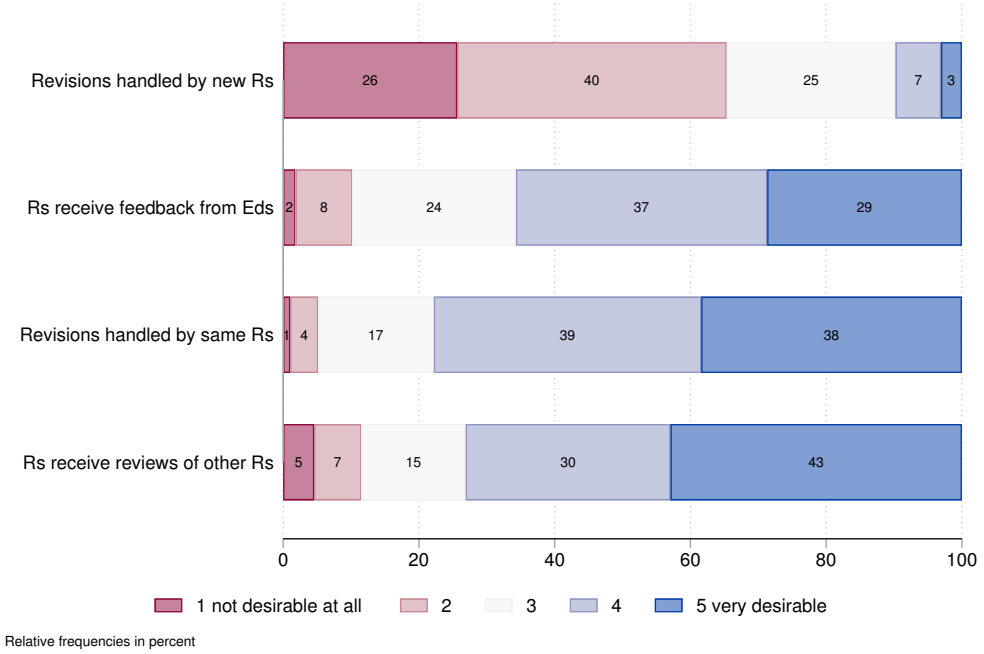


Figure 5: Desirable features of the review process (R denotes reviewer, $\bar{n} = 3,257$)

of the referees who were chosen to review the paper.[5] Figure 4 shows the degree of importance across the various items. It shows that the professional expertise of both reviewers and reviews are the most important criteria to assure best practice. Two thirds of respondents prefer a quick feedback in the form of a desk reject; a final decision about acceptance, revision or rejection within a month is also deemed desirable by the majority of scientists. Dimensions that appear to not serve as necessary conditions for the assurance of best practice are whether authors may propose reviewers, whether there is a compensation for reviewers or whether the review process is standardized in some way or another (e.g. by a short questionnaire).

Furthermore, we ask respondents about the standard procedures established within their respective disciplines. Interestingly, the single-blind review process appears to be more common than the double-blind process in some disciplines (see Figure A.1 in the Appendix). The reviews remain largely written as running text; bullet points or standardized ratings of articles via questionnaires are perceived as being less-established practices. Another set of questions asked what scientists perceive as desirable properties to assure best practice during the review process, especially with respect to their disciplinary background. Figure 5 illustrates that more scientists prefer the handling of revisions by the same reviewer(s) than the assignment of new referees. Respondents do not seem to care about the content of other reviewers if they themselves wrote one, while 55% consider feedback from the journal's editor as desirable or very desirable.

Finally, the last set of questions asked which properties of scientific journals are perceived to be the most significant for its quality. Figure 6 shows that – on the one hand – numerous special issues, the possibility that editors publish in their respective journal and a high degree of interdisciplinarity do not signal high quality. On the other hand, transparency about conflicts of interest, communication about the types of submissions handled by journals and the listing of journals in common repositories (e.g.

[5] Next to our own formulations, we also relied on templates assessing the quality of peer review procedures, foremost Wicherts (2016).

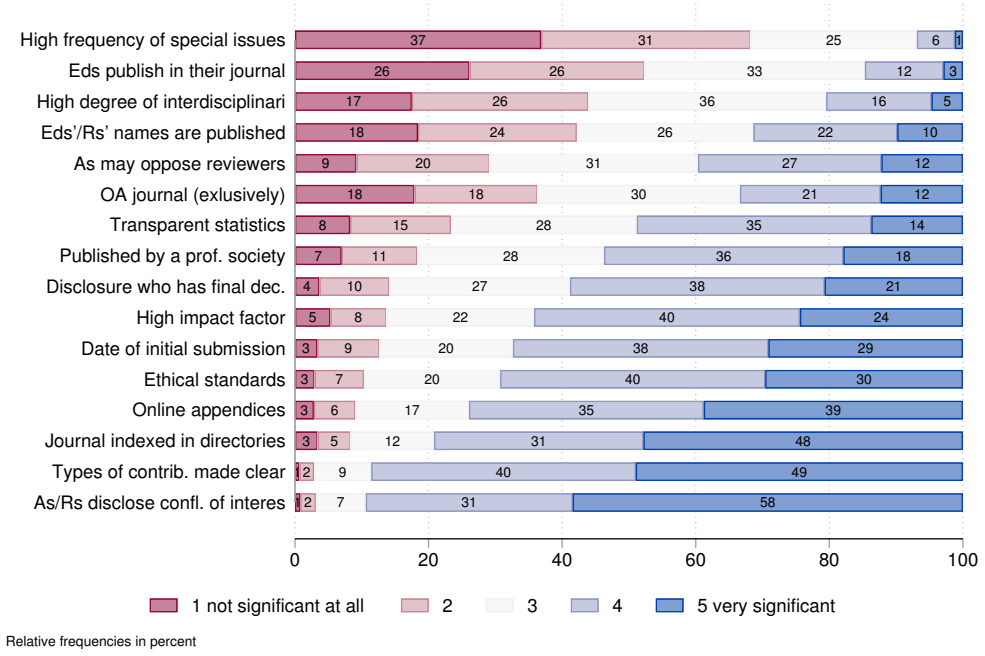


Figure 6: What contributes significantly to a journal's quality? ($\bar{n} = 3,254$)

PubMed, Scopus, Web of Science) are rated as important quality criteria.[6] Surprisingly, some properties reveal mixed significance, e.g. a high impact factor or that the journal is published by a professional society. The same holds true for whether a journal is exclusively published as an OA journal, which more than one third of respondents do not perceive being of utmost significance with regard to its quality.

While this descriptive analysis shows common trends in the answers across all respondents, our main interest lies in the variation between disciplines. For this reason, we apply the measure of the intraclass correlation coefficient (ICC) to assess the portion of the overall variance of single answers that can be attributed to some common characteristics. We group our analysis by disciplines. Two findings are noteworthy: First, we find that the disciplinary background of scientists does not explain any substantial variation across the eleven items regarding best practice. That is, expectations by scientists about the peer review process which can be considered best-practice are widely shared in the scientific community across all disciplines.[7]

Second, we were interested in which characteristics of a scientific journal scientists perceive as the most important with regard to the journal's quality within their respective disciplines. Again, we do not find substantial variation that can be attributed to the disciplines. Only a handful of items reveal that there are some differences between disciplines in comparison to the variation within the respective fields. On the one hand, single-blind or double-blind peer reviews represent process criteria that are judged differently across disciplines, at least to some extent. For instance, in the fields of humanities, economics or social sciences, double-blind reviews are perceived as the common denominator. In contrast, in fields like mathematics or pharmacy/biology, single-blind process appear to be much more common.[8]

[6] These results echo the efforts by Wicherts (2016), who puts a special emphasis on transparency in his investigation of peer review quality.

[7] Unsurprisingly, the same holds true for the expectations of scientists if answers are nested in institutions.

[8] See questions $Q2\_1$ and $Q2\_2$ in Figure A.3 in the Appendix, where we find the highest ICCs across all items that cover expectations.

There are two other areas where we find at least some variance attributable to the level of disciplines. One is the expectation towards a quick final decision (question $Q1\_10$); the other dimension covers whether journals should be indexed in relevant repositories and whether online appendices should be available on the journal's website ($Q4\_7$ and $Q4\_10$, see Figure A.5 in the Appendix).

4.2 Factorial Survey Experiments

Due to the complexities that may arise during review and decision procedures for journals and as an additional approach to assess scientists' expectations towards the quality of the review process, we conducted two factorial survey experiments. The goal was to investigate scientists' perception about the appropriateness and justification of specific decisions by (fictitious) editors and whether different properties of the review process are perceived as reputable. Factorial surveys – also known as vignette studies – consist of text descriptions where the content of the text is systematically varied across several dimensions by the survey researcher (Auspurg & Hinz 2014). Here, individuals read two or three vignettes per design and answered each one separately. That is, we applied two fully confounded factorial designs (Atzmüller & Steiner 2010), allowing for both the analysis of individual responses as well as the comparison of interaction effects between respondent-level and vignette-level characteristics.

4.2.1 Review Process

The first vignette study, denoted *review vignette*, describes the fictitious situation of an article submission by a team of authors for peer review at a scientific journal. The vignette text was varied with respect to six dimensions:

- Type of journal (open access journal // closed access journal)
- Anonymity of the review process (single-blind // double-blind)
- Selection process of reviewers (chosen by editor // chosen by the publisher via an algorithm)
- Number of reviewers (one reviewer // two or more reviewers)
- Decision options (either acceptance or proposal for resubmission // differentiated evaluation (i.e. accept, minor revision, major revision, reject))

One exemplary formulation was:

> "A team of authors submits a paper to an open access journal. The article is reviewed in a single-blind process by a reviewer selected by the editor. The reviewer has the option of accepting the article or proposing it for resubmission."

Each respondent read two vignettes that – in accordance to the fully confounded design – included all levels of the so-called vignette universe. Afterward, respondents were asked to answer on two 7-point rating scales how reputable they consider this journal to be (labeled "not serious at all" to "very serious") and whether they would submit an article to this journal ("very unlikely" – "very likely").

Figure 7 shows the results of two separate multilevel regression models on 6,502 answers i to the two vignettes nested within 3,266 respondents j.[9] From the magnitudes of the regression coefficients, we conclude that the property of having differentiated judgments about submitted articles and having at least two (or more) reviewers represent the two criteria that signal both high reputability and quality. Next to not having the opportunity to reject an article, it can be shown that the (anonymous) reviewer selection through an algorithm by the journal publisher (instead of the editor) has a detrimental effect on scientists' perception of quality. Two small findings are also interesting: First, OA journals are judged more positively by scientists within the vignettes than closed access journals. Second, despite the earlier statements that single-blind reviews represent a slightly more common practice in the respective disciplines, the vignette answers show that the double-blind peer review process is rated more positively, as it likely ensures impartiality to greater extent.

[9] Each vignette answer represents a single case. Because each respondent gives answers to two vignettes, the number of cases included in the analysis doubles. The deviation between number of respondents and number of given answers is due to item non-response.

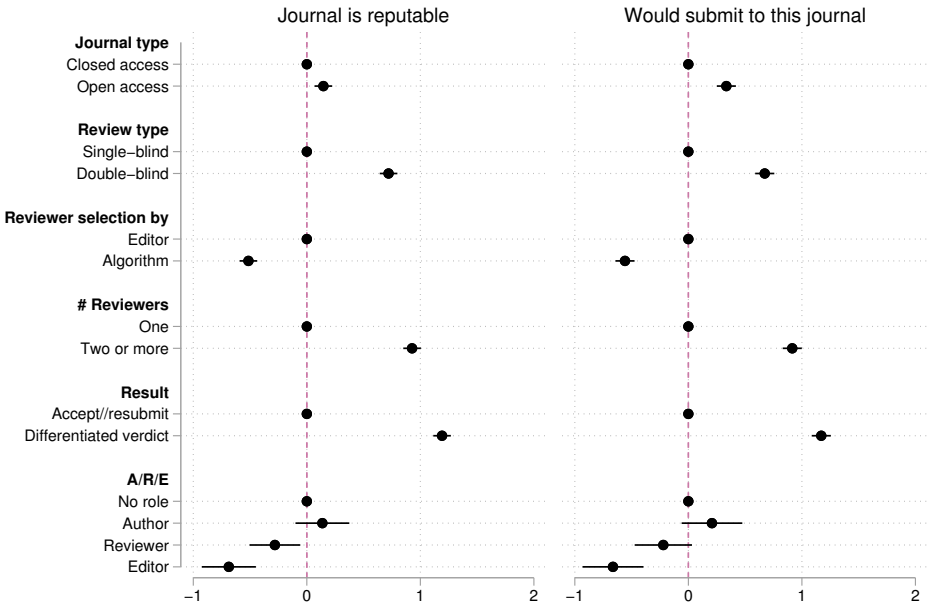


Figure 7: Estimated effects of properties of the review process on vignette answers about the reputation of fictitious journals (left panel) and about the probability of submitting to the described journal (right panel).

4.2.2 Decision Process

The second vignette experiment, called the *decision vignette*, has the objective to scrutinize decision practices by editors at journals. In particular, respondents were assigned to read three vignettes that covered different properties of the review process and features of the decision following the review process that was varied along the following dimensions:

- Duration of the review process (six months // two months // two weeks)
- Length of the reviews (each only about one page long // each about four to five pages long // each more than ten pages long)
- Recommendation of the reviewers (unanimous conditional acceptance // unanimous rejection // different decisions)
- Decision of the editor (reject the paper // accept the paper in its current state, regardless of the reviews // send the paper back to the authors for revision and resubmission)

As an example, one vignette text was as follows:

> "A team of authors has submitted an article to a journal. Six months after the initial submission, the authors receive two reviews, each about 4 to 5 pages long, with different recommendations concerning its acceptance. The editor then communicates his decision to send the paper back to the authors for revision and resubmission."

In contrast to the first vignette study, respondents now had to read three vignettes each and answer two follow-up questions after each vignette. This resulted in an overall sample size of 9,748 and 9,734 responses. We asked how they rated the quality of the review process and how they judged the editor's decision; all answers were documented on a 7-point scale with endpoints labeled "very bad" - "very good" and "unjustified" - "justified" respectively.

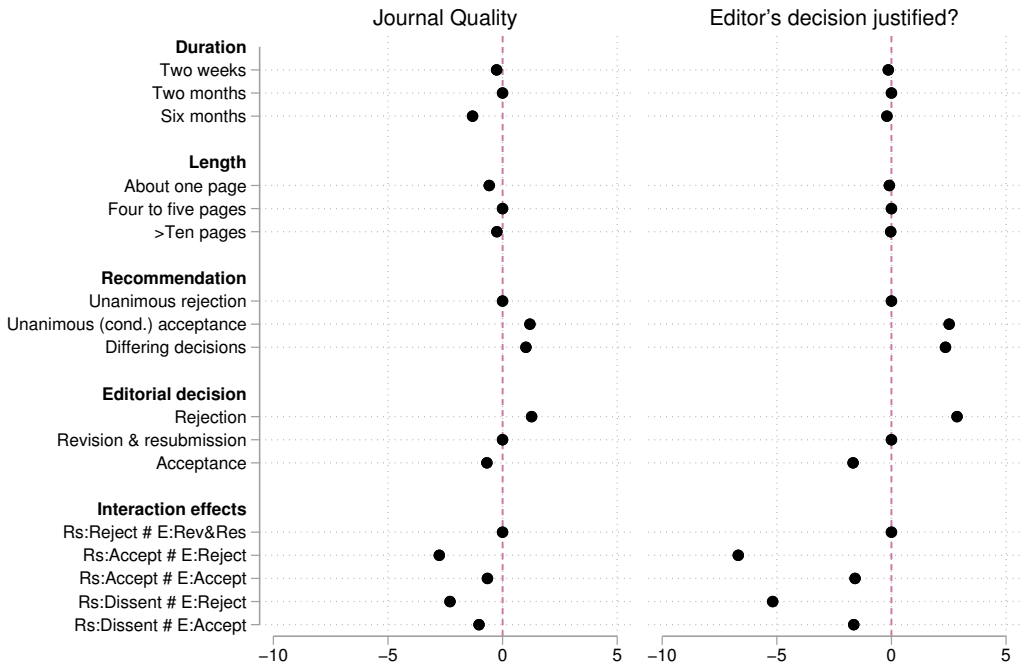


Figure 8: Estimated effects of properties of the decision process on the quality perceptions of respondents

We applied the same modeling approach (three vignettes nested within respondents) as earlier and report the results on the answers to the perceived quality of the journal in Figure 8. We find that respondents rate an extremely long (6 months) and a very short review duration as equally bad.[10] Very short reviews – with respect to length – likewise indicate poor journal quality. Furthermore, the judgment on the quality of a journal's decision process is shaped by the interplay of reviewer recommendations and the subsequent decision by the editor. Therefore, it is necessary to include interaction effects between the two dimensions in the model. The result of this procedure can be inspected in Figure A.6 in the Appendix, where we plot the perceived journal quality based on the editor's decision, conditional on the recommendations of the reviewers (x-axis).

The results also hold true for respondents' perception of the editor's decision as being justified. Here, process criteria such as duration and length of reviews do not play an important role in the perception. Instead, if decisions by the editor do not fall in line with recommendations by reviewers, decisions will be perceived as being unjustified, especially if acceptance is granted after a rejection by reviewers (see Figure A.7 in the Appendix). To summarize, poor editorial decision making in light of recommendations by reviewers (whether in agreement or in dissent) represents a key process criteria regarding the quality of a journal, even after properties of the review process are controlled for (duration, length).

[10] Therefore, we can replicate the finding by Huisman & Smits (2017) who also find lower quality ratings for review procedures that take a very long time. Unfortunately they did not investigate a potential curvlinear effect of review time on the quality rating.

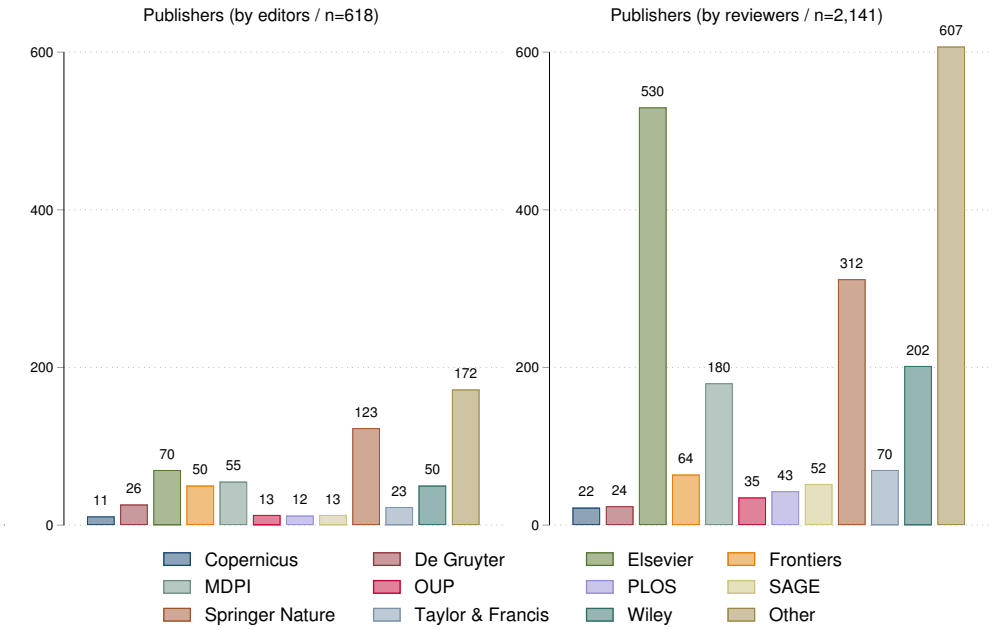


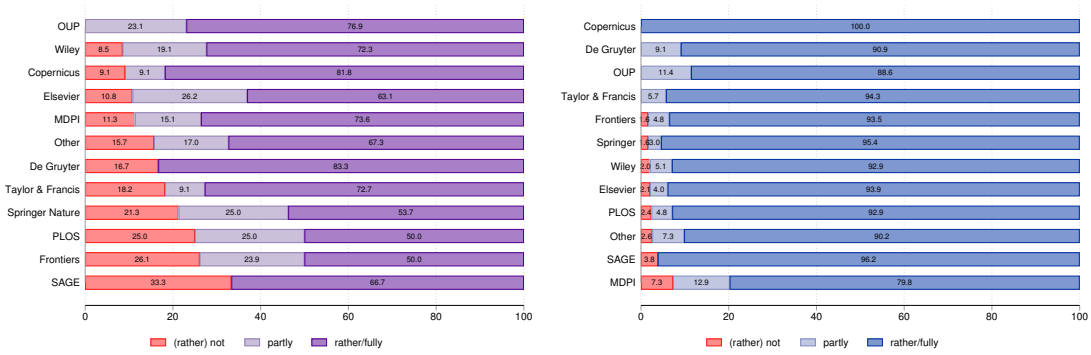
Figure 9: Absolute frequencies of publishers, as stated by editor's affiliation (left panel) and by reviewer's affiliation (right panel)

4.3 Assessment of Publisher's Quality Assurance

After the inspection of the expectations of scientists towards the quality assurance process across disciplines, we now focus on the actual experiences of scientists with different publishers. To reiterate, we found strong convergence of expressed expectations by scientists across disciplines, with only a handful of idiosyncrasies in certain disciplines (especially with regard to single-blind vs. double-blind as a standard). Due to the sequence of the questionnaire, we will present role-specific analyses, because the questions about the experiences with publishers during the review process differed between the thematic groups of editors, reviewers and authors. Our analysis across these three roles will focus on the quality dimensions that were rated as important to very important (see Figure 4). In particular, we will focus on the ways in which manuscripts are handled after the initial submission, on the quality and expertise of the review reports as well as the duration of the review process. The former two were mentioned as being very important to assure best practice; the latter played an important role in the decision vignette, where very short and excessively long duration times were signals for rather bad quality assurance.

4.3.1 Editors' perceptions

First, editors were asked which publisher their journal is published by. If they did not know the publisher, the journal name was asked and the publisher was later identified during the preparation of the data. All in all, we have data from $n_E = 618$ editors who indicated serving this role for a publisher's journal. The left panel in Figure 9 shows the absolute frequencies of stated publishers. First, we inspect whether manuscripts submitted to the editor(s) always fall within the actual scope of the editor's expertise. Figure 10a shows some amount of variation for the answers across publishers, with e.g. up to one third of all manuscripts submitted to SAGE journals to be considered as not matching the expertise of the respective editor. In contrast, editors for journals published by Copernicus or De Gruyter apparently were able to provide positive feedback with regard to their professional suitability.



(a) Editors' expertise ($N = 566$))

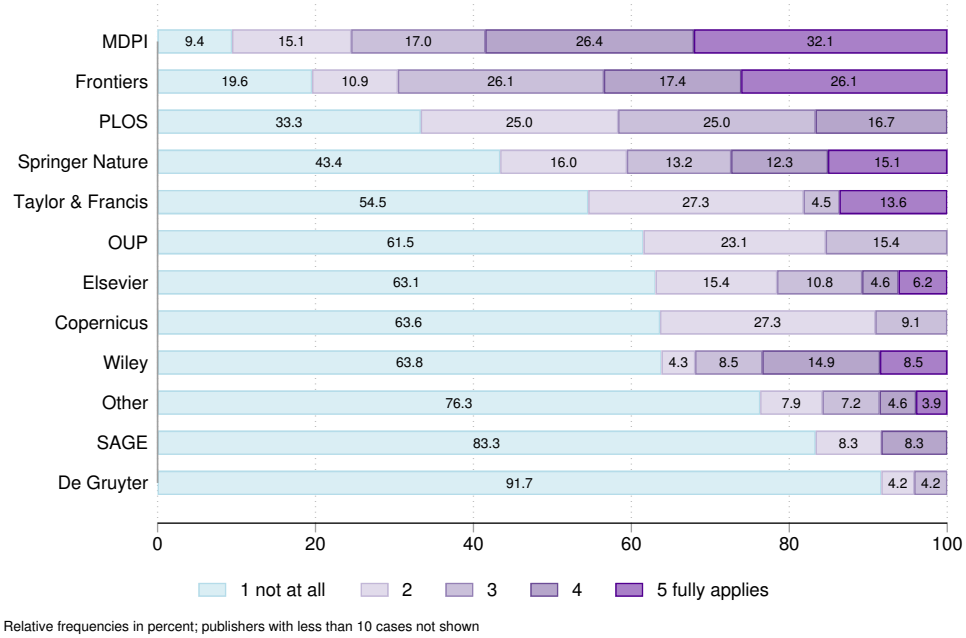
(b) Reviewers' expertise ($N = 2,103$)

Figure 10: Editors' and reviewers' opinions on their respective expertise concerning the subject of manuscript

Second, we consider how publishers might influence the dissemination of manuscripts to certain editors. Asked whether "manuscripts are assigned to editors by the publisher", we find substantial differences in the answers from editors, depending on whether they served for an OA publisher (e.g. MDPI, Frontiers) or other publishers (see Figure 11). For instance, whereas one third of the answers from editors who work for MDPI journals indicate that the manuscripts are disseminated by the publisher to the editors, this strategy does not seem to apply at rival publishers like Sage, Copernicus or De Gruyter. Of course, we have no information about the reasoning behind these different strategies by publishers.[11] We also asked whether reviewers are automatically selected and contacted by the publisher. Here we find evidence for the perception that OA publishers MDPI and Frontiers appear to be at the forefront of applying algorithms to select suitable reviewers (see Figure A.10 in the Appendix), a practice that respondents were rather opposed to and was perceived as a sign of low reputation in the review vignettes (see Figure 7).

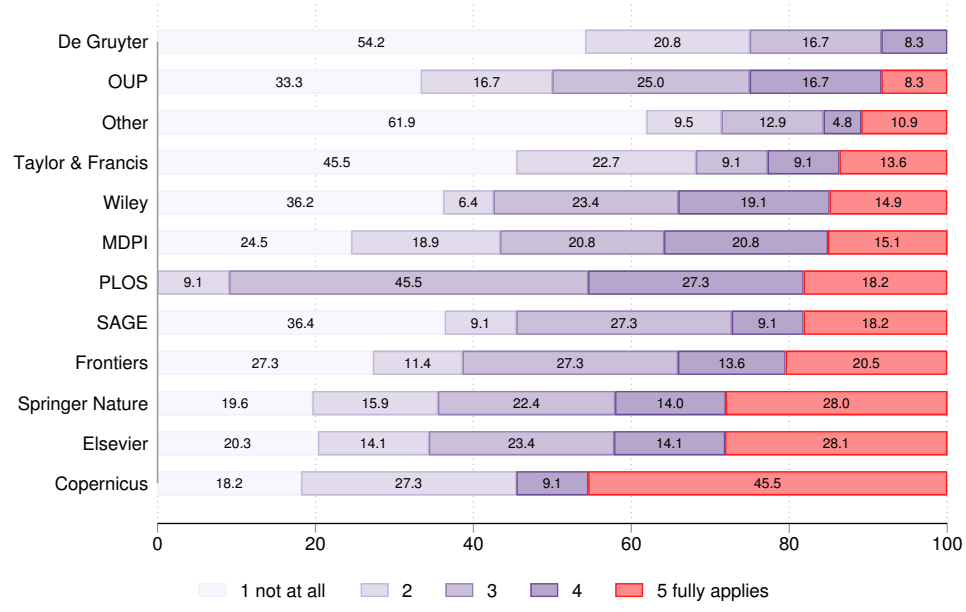
Additionally, we asked the editors whether they are allowed to forward manuscripts to journals of the same publisher, either before or after the review process. Figure 12 shows that this practice is actually quite common, also among the two largest publishers in academics, Elsevier and Springer Press.

[11] On the positive end of the spectrum, this practice streamlines the assignment of articles to editors given their expertise and accelerates the review process. On the negative end of the spectrum, publishers might play the role of gatekeeping – by funneling articles to particular editors with certain dispositions and preferences.



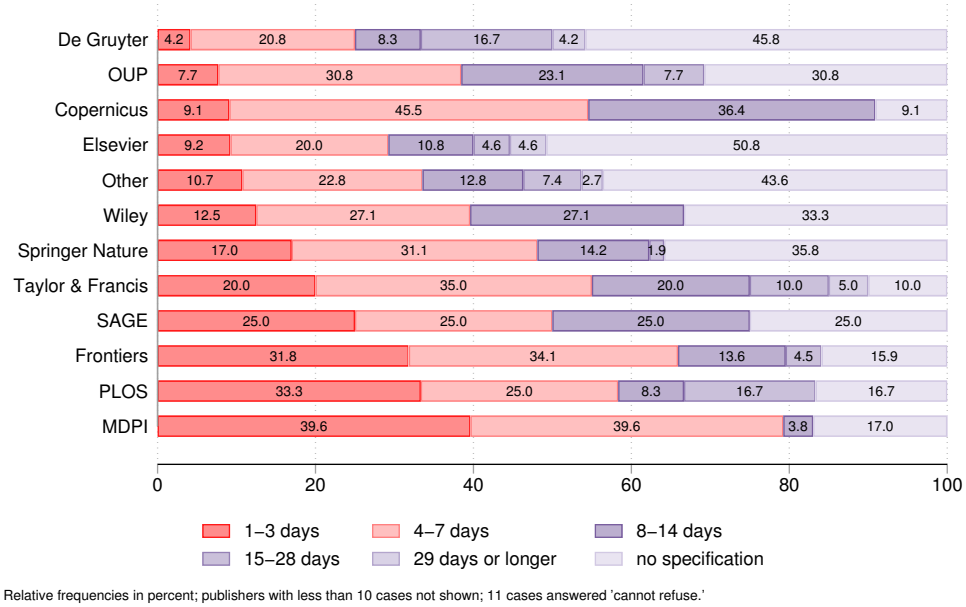
Relative frequencies in percent; publishers with less than 10 cases not shown

Figure 11: Are manuscripts assigned to editors by the publishers? ($n = 563$)



Relative frequencies in percent; publishers with less than 10 cases not shown

Figure 12: Are editors allowed to forward manuscripts to other journals by the same publisher? ($n = 553$)



Relative frequencies in percent; publishers with less than 10 cases not shown; 11 cases answered 'cannot refuse.'

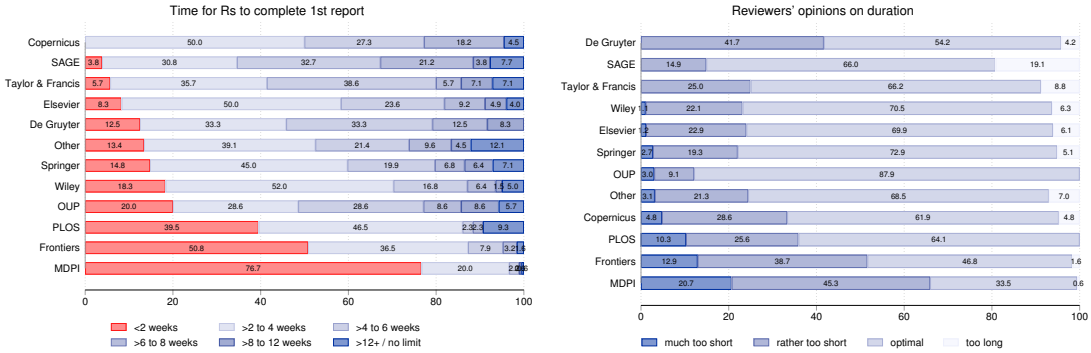
Figure 13: Time constraints on editors to arrive at first initial decision (e.g. desk reject)

Furthermore, we asked what time constraints are put on them to come up with a first initial decision to execute a desk reject or to assign reviewers. Figure 13 shows that most publishers do not put any specification in place with regard to the first initial decision, whereas MDPI, PLOS and Frontiers expect editors to handle submissions promptly and give feedback within a short time span. Answers about the time constraints for reviewers imposed by editors to submit their first referee report show that especially MDPI and Frontiers inflict time pressure on both reviewers to accelerate the review process, especially in comparison with other publishers (see Figure A.9 in the Appendix). The answers to the follow-up question whether editors consider this time constraint on reviewers as much too short, rather too short or optimal / too long clearly depict that editors are critical of MDPI, ahead of two other OA publishers, namely Frontiers and PLOS.

4.3.2 Reviewers' Perceptions

Similarly to editors, we first asked reviewers (and authors alike) for the name of the publisher for whom they reviewed an article in the last two month. Based on the assumption that reviewers and authors will most likely remember the journal name they reviewed for or submitted to rather than the name of the publisher, we asked for the name of the journal (via an auto-complete open text field) in case the publisher was unknown and later assigned the journals to their respective publishers. In total, we received $n_R = 2,141$ observations for the group of reviewers across the twelve publishers (including others). Figure 9 illustrates the distribution of absolute mentions of the publishers.

The most important expectation towards the review process was the professional suitability of the reviewers. We therefore asked whether reviewers considered themselves experts on the topic of the manuscript. In particular, we asked whether the review request fell within the scope of their professional expertise. Figure 10b shows that only one publisher slightly stands out as an outlier: Close to one fifth of the answers from reviewers who reviewed for MDPI journals stated that the manuscript did not fall or only partly fell into their areas of expertise. Similar to the answers by editors, it is stated by the reviewers that MDPI, Frontiers and PLOS ask reviewers most frequently to submit reviews within two



(a) Time constraints

(b) Opinions on duration

Figure 14: Rs' time constraints (left panel) and Rs' opinions on the duration of the review process, grouped by publisher.

weeks or less (see Figure 14a). Being asked about their opinion on the appropriateness of this duration, Figure 14b clearly shows that a majority of respondents considers this time span as being rather or much too short. Additionally, we asked reviewers whether they were pressured to finish their reviews by the respective journals. Figure 15 in the Appendix reveals that requests by publishers to complete the process is a property shared by all, but two publishers stand out. Frontiers and MDPI appear to pressure reviewers most often, with more than half of all respondents stating they were put on watch to fulfill their duties more or less explicitly.

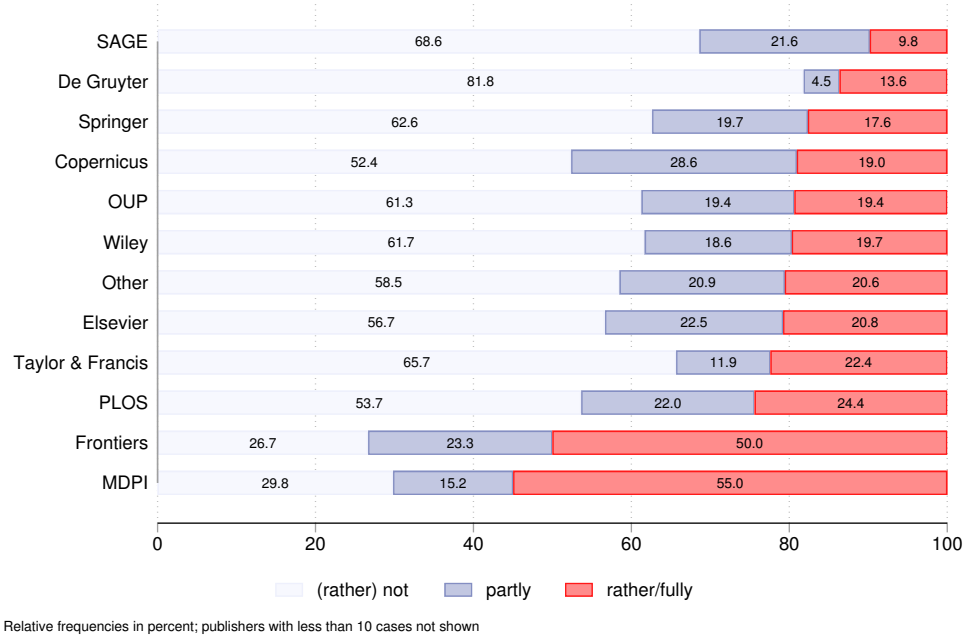


Figure 15: Pressure applied by publishers on reviewers to finish the review

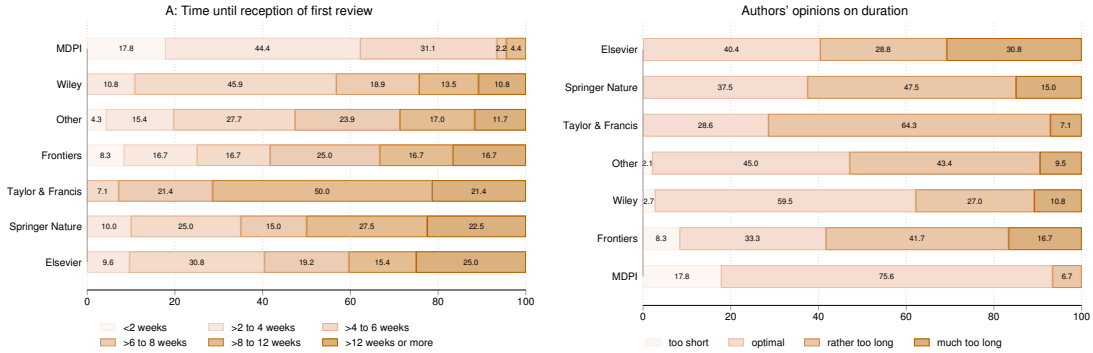
4.3.3 Authors' Perceptions

After the inspection of the expectations and experiences of editors and reviewers across publishers and whether they fall short on specific criteria of the quality assurance process, we now turn to the group of authors. Because most of our respondents indicated that they either served as an editor of a journal or reviewer of an article in the twelve months prior to the survey, the last group of authors who served in neither of these two functions represented the smallest group within the survey. Therefore, not all publishers could be covered in the analysis for this group. Figure A.8 in the Appendix shows the distribution of authors across the remaining publishers. The final sample size for the group authors was $n_A = 450$. Therefore, the caveat applies that our analysis of authors' perceptions rely on rather small sample sizes for individual publishers.

Being confronted with the question about whether reviewers always represent experts on the subject of the article, mixed results emerged (see Figure A.11 in the Appendix). Across all publishers, authors indicated that they were not entirely convinced that reviewers always represent experts on the subject of their manuscripts. Given that almost all authors may have reservations towards the opinions of reviewers and due to confidence in the validity of their own work, this more or less homogeneous result across publishers is not surprising.

Again, we find in Figure 16a that the OA publishers MDPI, PLOS and Frontiers allow authors to receive the first review after a rather short turnaround time. Concerning the first quality dimension of duration and in contrast to editors' and reviewers' opinions, authors rather perceive the short duration as being beneficial. The pattern shown in Figure 16b illustrates why there is demand for journals with short turnarounds and a high frequency of journal issues, as especially the short duration among MDPI journals is perceived as being optimal. Authors who feel the pressure of the highly competitive scientific job market ("publish or perish") simply have a preference for quick feedback on their articles. In contrast, scholars face the dilemma that short turnaround times put pressure on the editors and reviewers, while very short review times are perceived as a sign of low journal quality.

Finally, the positive judgment's of authors on quick decisions by certain journals may also be correlated



(a) Perceived duration

(b) Opinions on duration

Figure 16: Proportions of authors' answers on duration (left panel) and of their opinions about the duration until first review (right panel), grouped by publisher (in percent)

with high acceptance rates at these journals. As Huisman & Smits (2017: 645) have shown, "authors of accepted manuscripts give the peer review process a much higher rating than authors of a rejected manuscript." In fact, we also find this relationship in our data (results not shown): Authors who had their last submission rejected by the journal rate the quality of the review process (questions $f49\_1 - f49\_4$ in the Appendix B) substantially lower in comparison to authors of manuscript that were accepted.

5 Summary and Conclusions

The EQUAP[2] survey has provided key insights with regard to the three main goals of the project: First, with a few exceptions[12], *the expectations* towards the peer review process are very homogeneous across all disciplines. The most important criteria for the quality assurance are the transparency with regard to funding sources and the revealing of all possible conflicts of interest. Further, the expertise of the reviewers and the subsequent review process remains of utmost significance. We find considerable variance for the opinions on the importance of other quality criteria, with opportunities to propose reviewers and compensation for reviewers serving as dimensions that do not necessarily represent best practice. More than half of all scientists in the survey do not perceive a high frequency of special issues in a journal as a signal of high quality. The same holds true for the possibility of editors to publish in their own journal and when a journal's aims and scope are interdisciplinary. From the point of view of editors and reviewers, time represents an important quality feature, with very short and very long review times tending to be seen as a sign of poor publication practices. This finding is validated by the results of our vignette study on the judgment of journal quality.

Second, concerning the evaluation of the *experiences of scientists* – to what extent scientific publishers fulfill the expectations of researchers uniformly – we find more variation in the quality perceptions of respondents when stratified by journals. In particular, one finding was that selected publishers exhibit time pressure and assign very tight review times, as indicated by editors and reviewers. This especially holds true for publishers whose focus is on publishing exclusively OA. Another result was that some publishers tend to rely more heavily on automated assignments of manuscripts to editors and to reviewers.

[12]The notable difference was found in the expectations of scholars from the humanities and the social sciences (including economics) that peer review should be conducted as a double-blind process, whereas scholar from mathematics, pharmacy/biology and the natural sciences are more in favor of single-blind review procedures.

Based on our factorial survey experiment on the quality of the review process, we were able to show that such automatic assignments of reviewers based on an algorithm is perceived as less desirable by scientists, irrespective of their roles as editors, reviewers or authors. The question remains unsolved whether such automatic assignments are performed based on actual algorithms identifying potential reviewers due to thematic familiarity or whether it is steered by proposals from authors. The latter phenomena was investigated by Sarigöl et al. (2017), where the connectedness and shared biographies of authors and editors-in-chief resulted in quicker review times and higher rates of article acceptance in the OA journal PLoS ONE.

Third, the survey provides information that the perception and assurance of quality in peer review processes is subject to various conflicts of interest. Criticism of various practices of the publication process by editors and reviewers is met with much less criticism by authors. Authors face the challenge of long waiting times for receiving their feedback while being pressured by temporary employment and by the mantra "publish or perish" (Harzing 2010). Hence, it is unsurprising that quick review times are not perceived as critical as by other stakeholders in the review process.

One conclusion we draw from this survey is that commercial OA publishers (e.g. MDPI, Frontiers or PLOS) represent beneficiaries of the external effects that are evoked by the peer review process of scholarly publishing. On the one hand, editors and reviewers alike contribute voluntarily to the review process while facing ample times constraints, dealing with problems like finding suitable reviewers and handling submissions and reviews along with their everyday professional lives. On the other hand, authors – especially in their predoc or postdoc roles – have a desire to receive quick and comprehensive feedback on their submissions because published papers represent the foundation for funding proposals and a successful academic career. As our results show, quick decisions are viewed as being desirable from the author's perspective (see also Huisman & Smits 2017).

Some publishers have identified this demand for quick processing of the review and publication process and have established an incentive structure to ensure quick turnaround times by introducing procedures of automatic reviewer assignment, pressuring editors and reviewers alike to fulfill their duties and by increasing the number of special issues. Some scholars have coined this strategy "aggressive rent extracting" (Crosetto 2021). But respondents also stated that they perceive very short review times as a sign of low journal quality, while often being pressured to finish a review in a timely fashion.[13] If scholars excessively opt to publish their articles in questionable outlets, the reputation mechanisms will create a negative feedback loop. Hence, such signs of low research quality within individual CVs will adversely affect the careers of these scholars. Still, publishers and professional societies alike face the challenge that the current quality assurance process within peer reviewed journals is marked by severe conflicts of interest. Given the entrepreneurial motivation of private corporations to preserve the status quo, true and honest reconciliation of these conflicts has to come from within academia (Aspesi & Brand 2020).

[13] This results was derived from our vignette experiment as well as from the role-specific answers of our respondents.

References

Aspesi, Claudio & Amy Brand (2020). In pursuit of open science, open access is not enough. *Science* 368.6491, pp. 574–577. DOI: 10.1126/science.aba3763.

Atzmüller, Christiane & Peter M Steiner (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6.3, p. 128. DOI: 10.1027/1614-2241/a000014.

Auspurg, Katrin & Thomas Hinz (2014). *Factorial survey experiments*. Vol. 175. Sage Publications. DOI: 10.4135/9781483398075.

Black, Nick et al. (1998). What makes a good reviewer and a good review for a general medical journal? *JAMA - Journal of the American Medical Association* 280.3, pp. 231–233. DOI: 10.1001/jama.280.3.231.

Crosetto, Paolo (2021). Is MDPI a predatory publisher? *URL: https://paolocrosetto. wordpress. com/2021/04/12/is-mdpi-a-predatory-publisher (accessed: 19. 11. 2022.)*

Harzing, Anne-Wil (2010). *The publish or perish book*. Tarma Software Research Pty Limited Melbourne, Australia.

Huisman, Janine & Jeroen Smits (2017). Duration and quality of the peer review process: the author's perspective. *Scientometrics* 113.1, pp. 633–650. DOI: 10.1007/s11192-017-2310-5.

Sarigöl, Emre et al. (2017). Quantifying the effect of editor–author relations on manuscript handling times. *Scientometrics* 113.1, pp. 609–631. DOI: 10.1007/s11192-017-2309-y.

Stadtmüller, Sven et al. (2019). Adaptation of the AAPOR final disposition codes for the German survey context. *Mannheim, GESIS–Leibniz-Institute for the Social Sciences (GESIS–Survey Guidelines)*. 10.

Weber, Ellen J et al. (2002). Author perception of peer review: impact of review quality and acceptance on satisfaction. *JAMA* 287.21, pp. 2790–2793. DOI: 10.1001/jama.287.21.2790.

Wicherts, Jelte M (2016). Peer review quality and transparency of the peer-review process in open access and subscription journals. *PloS ONE* 11.1, e0147913. DOI: 10.1371/journal.pone.0147913.

Appendix A

Sample

Table A.1: Sampling information from the 25 participating institutions

#	Institution	Abbrev.	Invitees	Resp.	Mode of invitation
1	Karlsruhe Institute of Technology	KIT	.	16	unknown
2	Leibniz University Hannover/TIB	LUH/TIB	.	187	mailing list
3	RWTH Aachen	RWTH	.	283	mailing list
4	TU Berlin	TUB	.	72	unknown
5	TU Braunschweig	TUBR	3,500	187	total population
6	TU Darmstadt	TUDA	2,951	224	total population
7	TU Dresden	TUD	7,534	338	total population
8	TU München	TUM	13,682	733	total population
9	University Stuttgart	UST	.	159	unknown
10	University Basel	UNIBAS	.	3	mailing list
11	University Bern	UNIBE	.	7	mailing list
12	University Zürich	UZH	5,094	271	total population
13	Université Geneva	UG	.	4	newsletter
14	ZHAW Zurich University of Applied Sciences	ZHAW	.	3	mailing list
15	Université de Neuchatel	UNINE	.	17	mailing list
16	Université de Lausanne	UNIL	.	4	unknown
17	École Polytechnique Fédérale de Lausanne	EPFL	2,854	152	total population
18	University of Applied Sciences & Arts Northwestern Switzerland	FHNW	.	10	mailing list
19	Université Fribourg	UNIFR	.	138	total population
20	University of Applied Sciences Dresden	HTWD	.	20	mailing list
21	FZ Jülich	FZJ	.	91	mailing list
22	GESIS Leibniz Institute for the Social Sciences	GESIS	.	46	total population
23	Leibniz Institute for Prevention Research & Epidemiology	BIPS	.	18	mailing list
24	University of Tübingen	TUEB	8,637	222	total population
25	TU Bergakademie Freiberg	TUFR	.	65	mailing list

Table A.2: Socio-demographic information – respondents' gender

Category	Absolute frequencies	Relative frequencies	Cumul.
Female	933	29.60	29.60
Male	2,126	67.45	97.05
Diverse	32	1.02	98.06
Nonresponse	61	1.94	100.00
Total	3,152	100.00	

Table A.3: Socio-demographic information – respondents' academic position

Category	Absolute frequencies	Relative frequencies	Cumul.
Research Assistant (predoc)	1,047	32.02	32.02
Research Assistant (postdoc)	1,002	30.64	62.66
A13 / Junior Prof. / Emeritus	344	10.52	73.18
Full / Research Professor	584	17.86	91.04
Others / Missing	293	8.96	100.00
Total	3,270	100.00	

Expectations and Common Standards

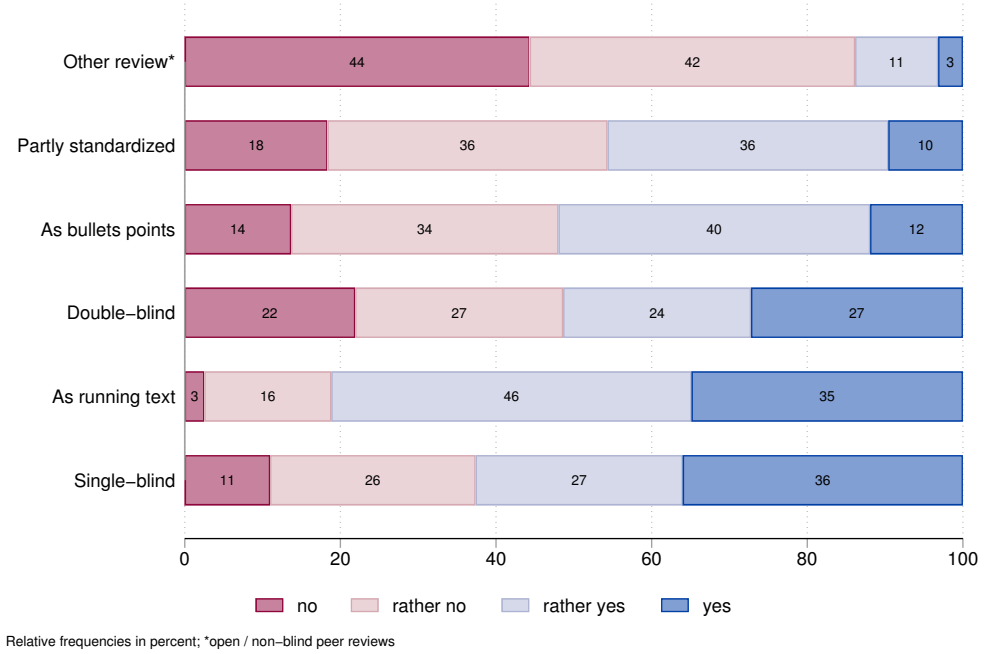


Figure A.1: Common standard within respective disciplines ($\bar{n} = 3,180$).

Expectations Across Disciplines

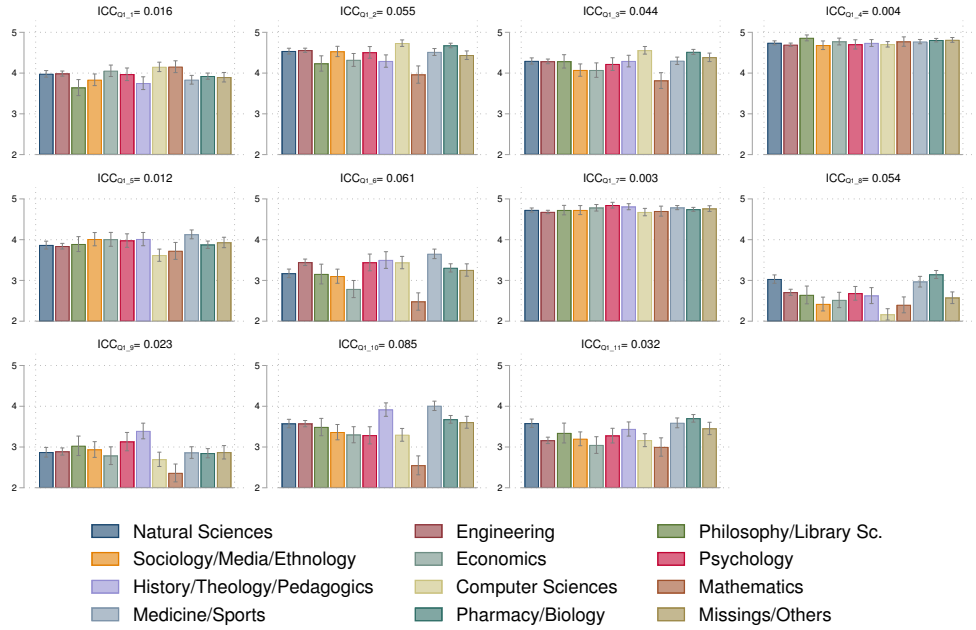


Figure A.2: Mean expectations towards best practice – answers grouped by disciplines*

*Indices $Q1_1$-$Q1_{11}$ refer to the block of questions denoted f1_1-f1_11 in Appendix B. The same logic applies to the following plots in Figures A.3 to A.5.

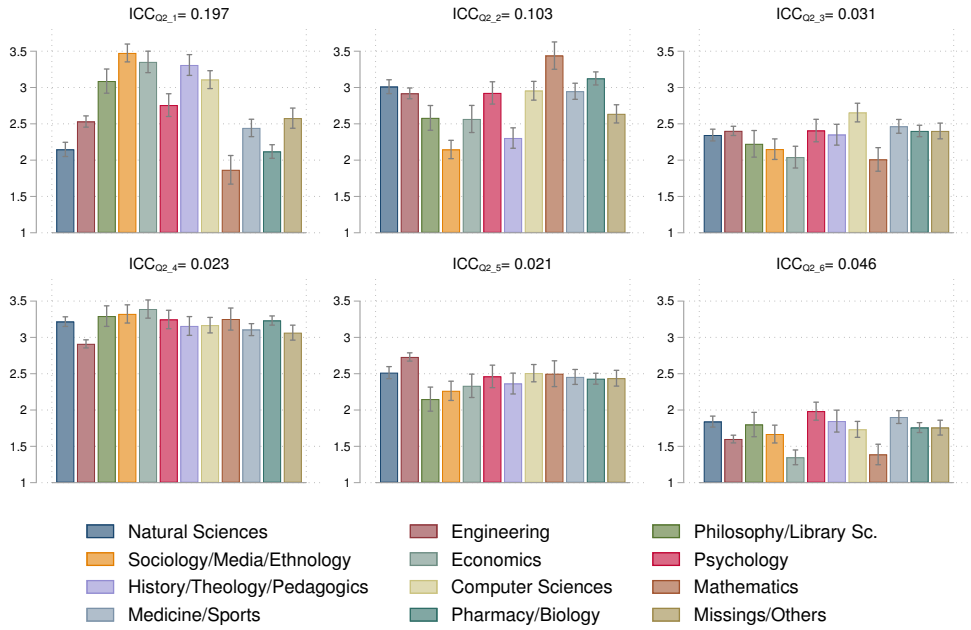


Figure A.3: Mean expectations towards best practice – answers stratified by disciplines

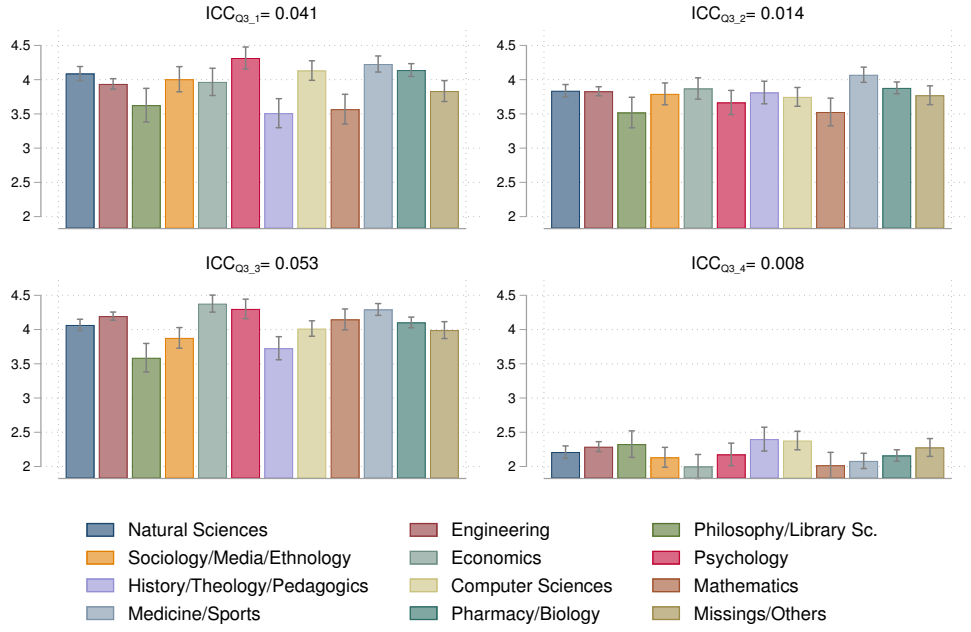


Figure A.4: Mean expectations towards best practice – answers grouped by disciplines

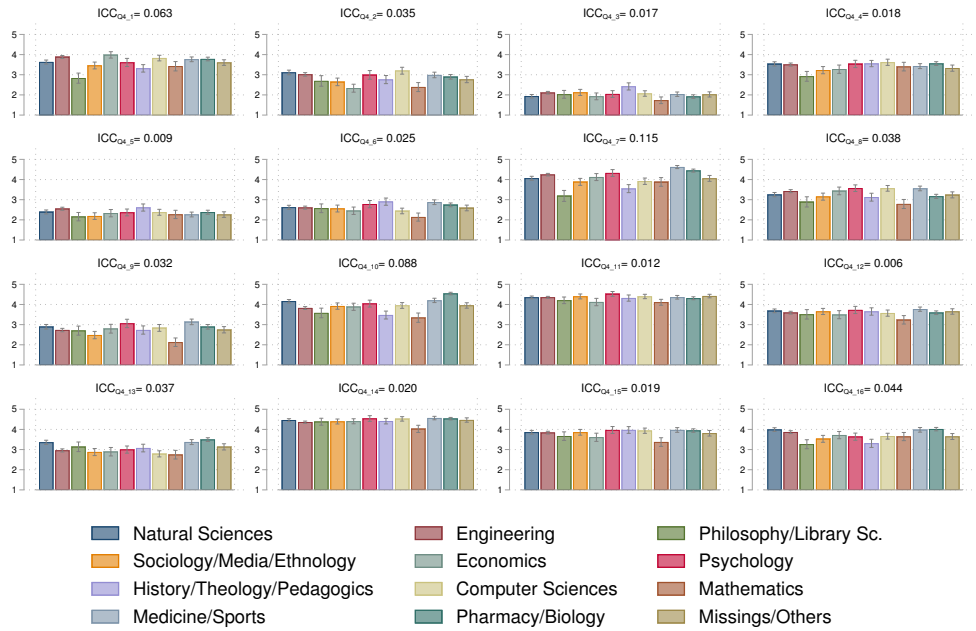


Figure A.5: Mean expectations towards best practice – answers grouped by disciplines

Vignette Experiments

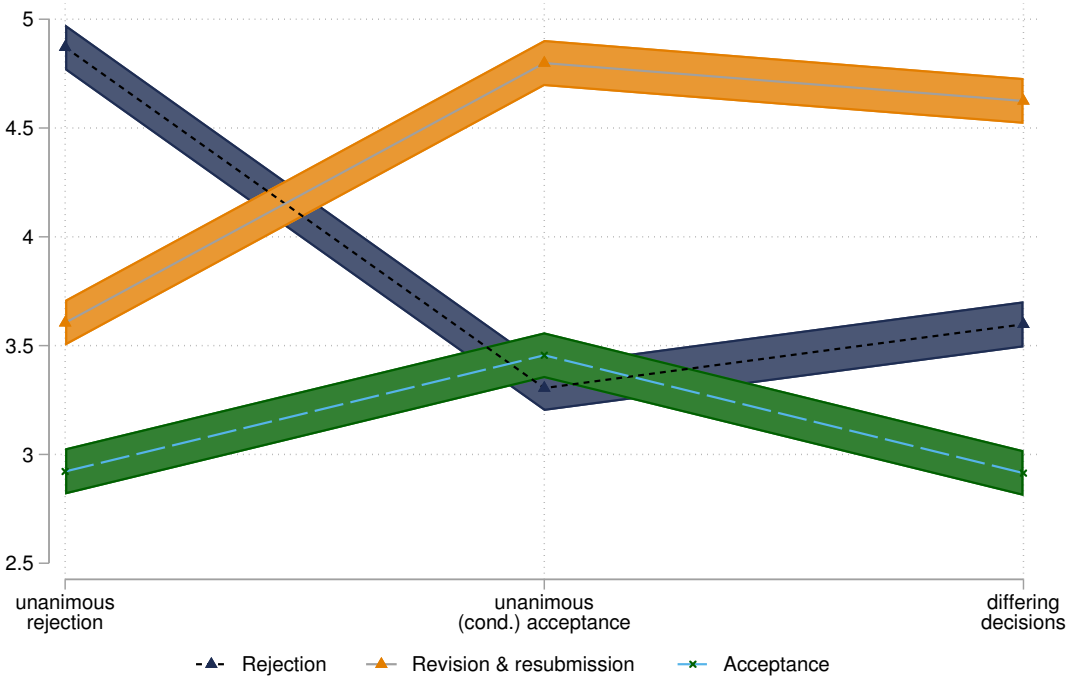


Figure A.6: Marginal effects of editor decisions conditional to reviewer recommendations on the perceived journal quality

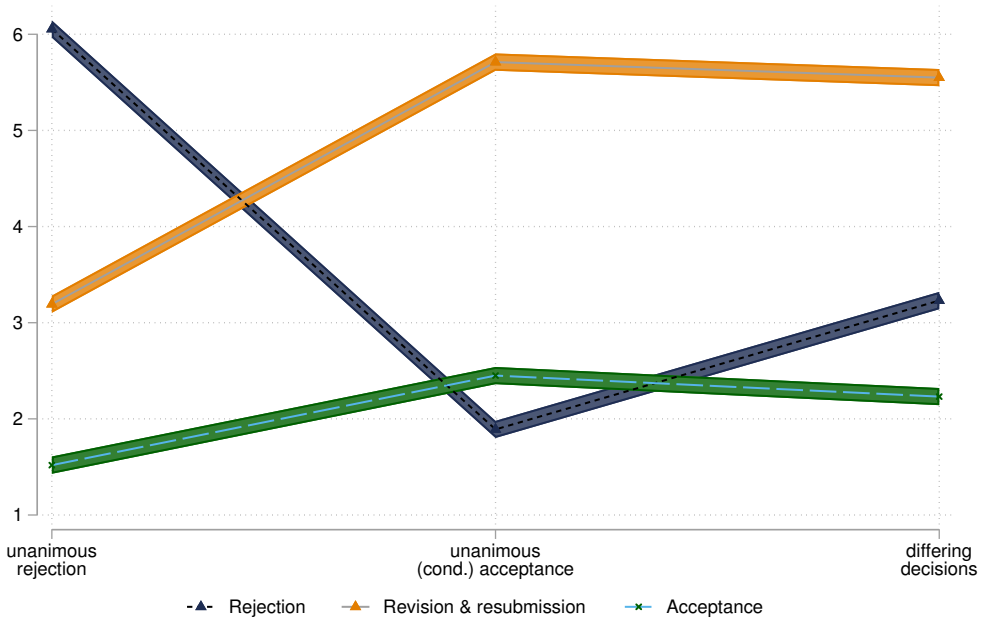


Figure A.7: Marginal effects on the justification of the editor's decisions, conditional to reviewer recommendations

Perception Across Publishers

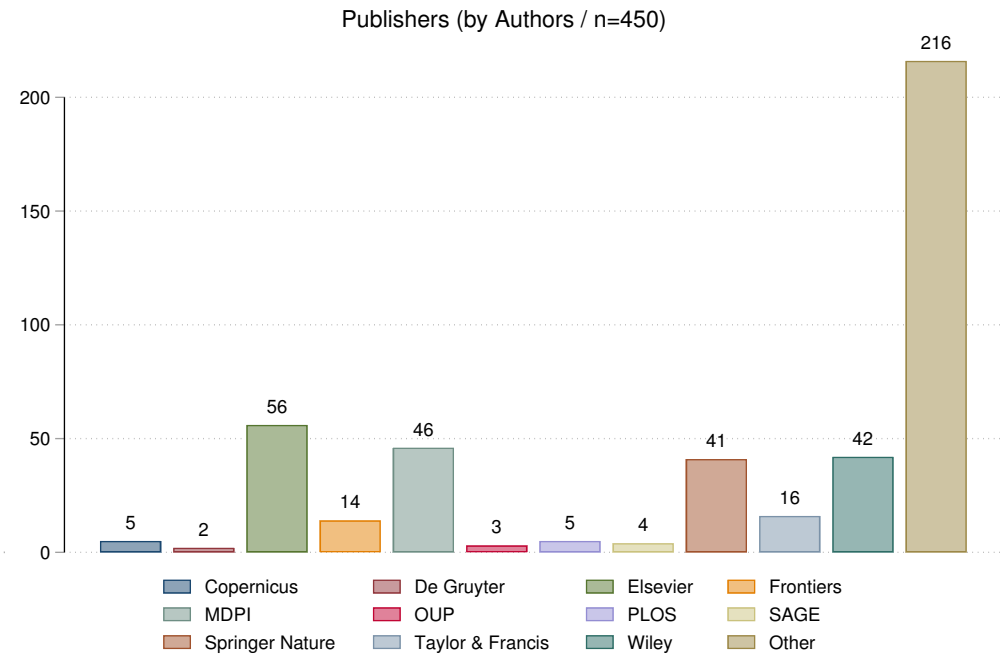


Figure A.8: Absolute frequencies of publishers, as stated by authors' answers to which journal they last submitted

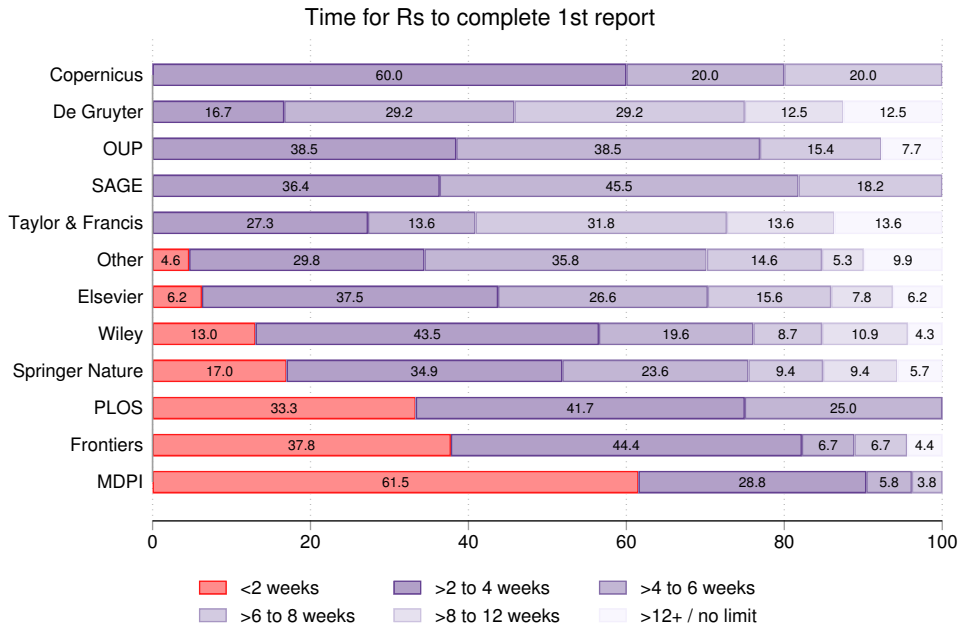


Figure A.9: Time constraints put on reviewers to complete first review report, as indicated by editors

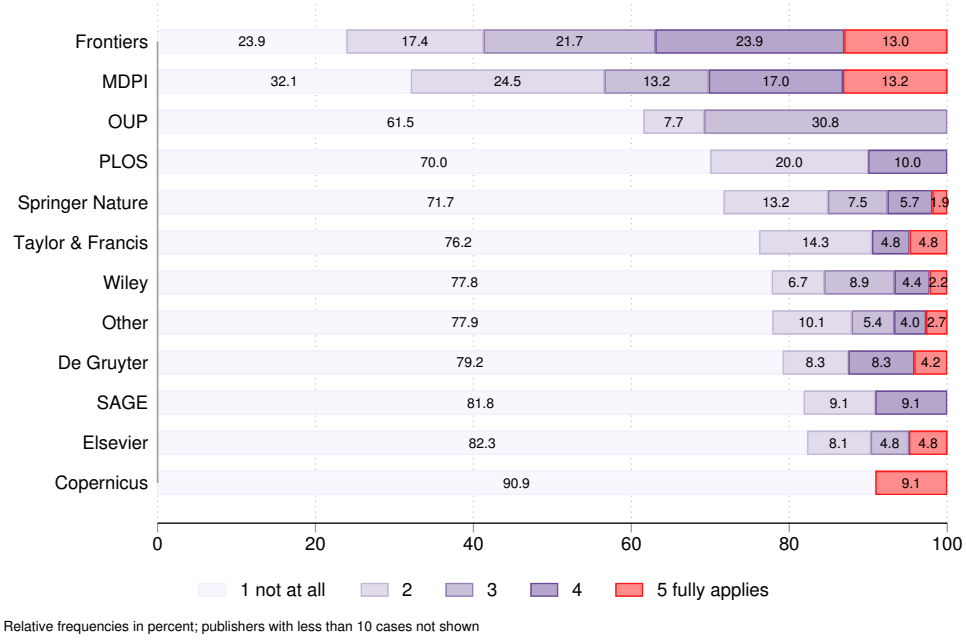


Figure A.10: Are reviewers automatically selected and contacted by publishers, as answered by editors

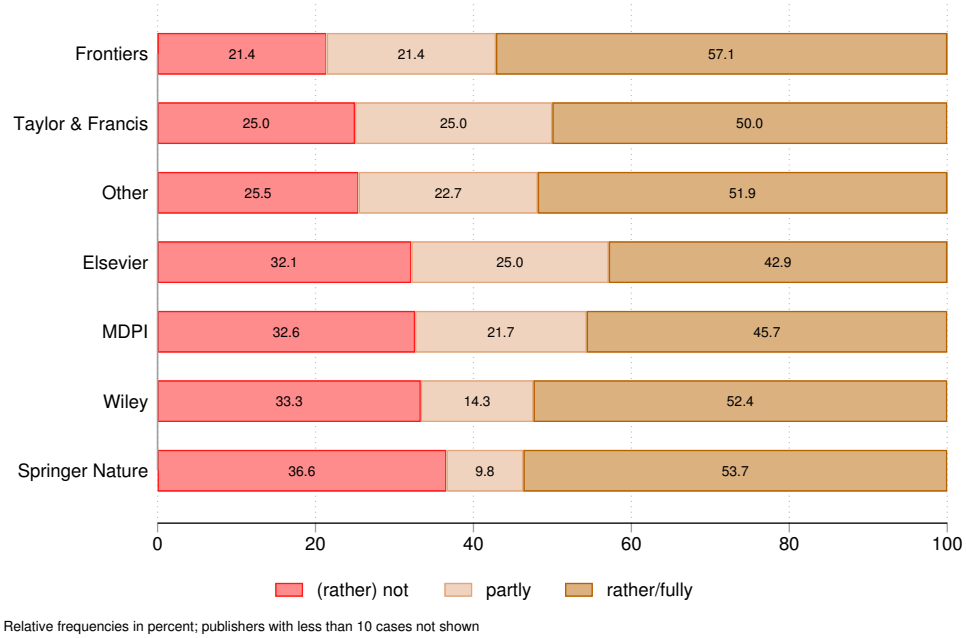


Figure A.11: Authors' opinions on "Reviewers are always experts on the subject of the article", grouped by publisher