How do journals of different rank instruct reviewers? An analysis of reviewer guidelines in the field of management.

Marco Seeber[1]

[1] Department of Political Science and Management, University of Agder, Universitetveien 25,

6404 Kristiansand, Norway

Keywords

Peer review; reviewer guidelines; sociology of science; economics of science; management science; status; ranking; validity; novelty; gatekeeping; developmental.

Abstract

Current knowledge on peer review relies on general formulations of its goals and micro level accounts of its practice, whereas the role played by journals in shaping peer review has been hardly investigated so far. This article addresses this gap by studying the reviewer guidelines (RG) of 46 journals in the field of management, by focusing on variations in instructions for reviewers between journals of different status rank. While evaluation and assessment procedures rely on formalized criteria and guidelines, this is not common practice in scholarly peer review. However, reviewers are prone to use heuristics that may have unintended effects. Thus, journals may adopt RG to nudge reviewers using proper evaluation criteria and consistently with the journal's needs.

The analysis reveals comparative differences between RG of high and low rank journals. RG of high rank journals are texts, which stress the theoretical contribution and methodological validity in strict terms. The RG of low rank journals are mostly forms, which give more emphasis to the empirical contribution and the quality of communication. RGs of very high rank journals stand apart when compared to other journals, as they include 45% less gatekeeping instructions but four times more developmental instructions, arguably to enhance the capability to detect the most innovative contributions. However, the fact that developmental instructions decrease from top to low rank journals may represent another case of cumulative advantage in science.

1. Introduction

Peer review aims to guarantee the quality of the research published and to provide credibility to scientific claims (Warren 2003; Kalleberg 2012; Starbuck 2003; Bornmann 2008). Therefore, peer review is adopted by nearly all academic outlets as the legitimate way to select scientific contributions (Lamont 2009), and it is the key process generating publication traces that are at the heart of bibliometric studies (Squazzoni et al. 2017).

While peer review is a highly institutionalized practice, yet several studies point out that reviewers' and editors' behaviours are affected by non-scientific elements as well, which question the impartiality and reliability of peer review (Eysenk and Eysenck 1992; Campanario 1998a,1998b). These works unveiled the nature of peer review as a social process (Knorr-Cetina 1981; Ellison and Rosato 2002), affected by the relationships between editors, authors and reviewers, the setting of incentives, as well as authors' demographic traits, country of origin, institutional affiliation, and past productivity (e.g. Sandström and Hällsten 2007; Lee et al. 2013; Balietti et al. 2016; Sarigöl et al. 2017; Seeber and Bacchelli 2017;).

A remarkable gap in research on peer review is that variations between different journals have been largely ignored so far. Descriptions of peer review as a practice, its goals and principles of functioning, are rather general (Weller 2001): reviewers are expected to act as gatekeepers, namely to select the best papers in terms of their contribution, validity and clarity (Hackett and Chubin 2003), to spot errors and scientific misconduct (Smith 2006), to provide suggestions for improvement (Schiminke 2002). On the other hand, with few

exceptions, (e.g. Siler and Strang 2017) analyses of micro level functioning of peer review hardly took into consideration the characteristics of the field and the journal.

This gap may be due to the fact that, while in most evaluation settings making criteria explicit and providing guidelines to experts is very common– such as in project selection or recruitment – this is not the case for scholarly peer review, where reviewers are left free to interpret their role. Rennie (2015) argued that a reverence for peer review bordering on mysticism has prevented research on how to improve peer review, while guidelines can be helpful indeed, for example, research on clinical reporting led to develop 'risk of bias' assessments to help reviewers to evaluate possible weaknesses in trial reports.

General and micro level perspectives can be complemented by a meso level perspective, focusing on the active role of journals in guiding peer review. In times when peer review is under fire (McCook 2006), a focus on journals sheds light on how they try to remedy peer review limitations. Moreover, variations in how journals instruct reviewers can have important consequences for knowledge production and reputational dynamics in science. Practices are in fact adapted to local contexts according to specific organizational needs (Czarniawska and Joeges 1996), and given that journals have very different needs and resources, considerable differences are expected in how they try to shape peer review.

Therefore, the goal of this article is to explore how journals instruct reviewers and particularly variations between journals of different rank. We analyse the instructions for reviewers in the reviewer guidelines (RG) of 46 *management* journals included in the Web of Science database (WOS). RGs deserve close inspection as they are journals' official documents, potentially important in shaping peer review.

The following section discusses the role RG can play in addressing some key challenges of peer review, and develop propositions on how journals' rank relates to instructions for reviewers. The data and methods of analysis are presented in the third section, while the empirical analysis is presented in section four. The final section discusses the article's main findings and their implications for peer review and knowledge production.

2. Theoretical framework

2.1 Peer review: a blanket too short?

Peer review fulfils a *gatekeeping* function in selecting the best contributions, with authors making knowledge claims and reviewers assessing their *validity* – e.g. by scrutinising errors, uncovering misconduct, seeking inconsistencies, - and their *novelty*/importance (Zuckerman and Merton 1971; Ziman 1984; Latour 1987; Bedeian 2004; Smith 2006).

The validity and novelty of knowledge claims affect the legitimacy and reputation of scientists and journals alike (Merton 1968; Patriotta 2017). However, the capability of peer review to assess a contribution's validity and novelty is far from perfect.

The *condicio sine qua non* for *validity* is that data and results have not been manipulated (Bhattacharjee 2013). Yet peer review has limited capability to detect similar scientific misconducts (Haug 2015). Another notable challenge is related to the so-called publication bias, namely that an article's chances to be published are higher when presenting surprising and large effects, which are revealed by later studies to be not large or even null (Dickersin 1990). According to Brehms et al. (2013), such issue is particularly salient for high rank journals, and likely responsible for the worrisome negative relationship often observed between journals' rank and the validity of the published findings.

The capability of peer review to identify innovative contributions is limited, among others, due to the fact that the talent to evaluate what are the fundamental problems is a rare gift (see Zuckerman 1977 on Nobel prizes winners' "taste for science"), that reviewers acting for top-tier journals may put the selectivity bar very high

as a way to impress the editors while losing sight of the importance of the contribution (Romanelli 1996),[1] and that contributions that can disrupt current understandings are likely to undergo more severe scrutiny (Kuhn 1962, 1977; Moran 1998; Siler and Strang 2017). As a matter of fact, many breakthrough contributions later awarded Nobel prizes have undergone particularly lengthy and troublesome peer review processes (Campanario 1996; 2009). Siler and Strang (2017) examined the review process of 52 scholars who published in Administrative Science Quarterly journal and found a rather conservative attitude in peer review, and papers challenging established theoretical perspectives faced distinctively high levels of criticism. Siler et al. (2015) found out that three elite medical journals rejected many high impact articles, including the 14 most cited ones – roughly the top 2%.

At journal level, only a weak positive relationship exists between journal rank and novelty-importance of contributions (Allen et al. 2009; Evangelou et al. 2012), while Chen and Konstan (2010) found that selectivity and impact of conferences in computer science correlate up to an acceptance rate of 20%, yet conferences with an acceptance rate of 10-15% have significantly lower impact than conferences with a 15%–20% acceptance rate.

In sum, peer review seems to face some trade-offs. An emphasis on selectivity and validity may increase the risk of rejecting impactful works (Miner 2003; Patterson 2004; Starbuck 2003), whereas an emphasis on novelty may further enhance a publication bias, endangering the reliability of the published works (Brehms et al. 2013).

2.2 Instructions for reviewers

Scientific journals compete for the attention and approval of the scientific community. It is then crucial for journals to select valuable manuscripts, and in order to do that they may attempt to guide reviewers' behaviour. A major factor that is likely to affect journals' instructions to reviewers is their rank, since ranking position affect a journal's resources and objectives. The following paragraph develops four hypotheses on the relationship between a journal's rank and the instructions for reviewers.

2.2.1 Journal rank and instructions for reviewers

Journals improve their reputation and ranking position by identifying impactful publications. Given some conservative and risk-minimizing aspects of peer review (Langfeldt 2006), stressing the gatekeeping function may be counterproductive and lead to the rejection of innovative contributions. In recent years, scholars have highlighted that beyond a gatekeeping function, peer review also serves a *developmental* function in improving manuscripts (Schiminke 2002). While a developmental approach to peer review is not devoid of potential downsides (Bedeian 2004),[2] on the other hand developmental instructions, such as to provide suggestions and contribute recovering a manuscript from its limitations, may reduce the risk of "throwing the baby out with the bathwater" and hence of innovative contributions being rejected.

A developmental peer review is however highly demanding to reviewers, and arguably not all journals can realistically ask reviewers to adopt a developmental approach. Organizational research highlights that high status organizations can extract greater effort from their partners, because collaboration with a high status organization generates positive returns in terms of reputation and prestige (Castellucci and Ertug 2010). In a similar vein, being a reviewer for a high rank journal may have greater returns, meaning these journals are in a stronger position to give developmental instructions for reviewers. Hence:

[1] Also known as SLAMing – i.e. Stressing the Limiting Aspects of a Manuscript (Bedeian 2004).

[2] Such as to violate authorial voice, by affecting a manuscript to the extent that it does not anymore represent authors' intentions, goals and style (Bedeian 2004).

Hypothesis 1: RGs of high rank journals include a greater number of developmental instructions than RGs of non-high rank journals

Theory building plays a central role in scholarly communication (Hambrick 2007), and are particularly salient for top-tier management journals (Corley and Gioia 2011), which pursue a positivist agenda concerned with identifying general rules/theories (Legge 2001). Therefore:

Hypothesis 2: Instructions to assess a manuscript's theoretical contribution are more common in RGs of high rank journals

Practical contributions are also very valuable in the field of management (Corley and Gioia 2011), yet given the prominent role of top-tier journals in theory building, they are likely to be more important for non-high rank journals.

Hypothesis 3: Instructions to assess a manuscript's practical contribution are more common in RGs of non-high rank journals

Different expectations in terms of novelty arguably imply different requirements in terms of validity. Namely, a practical contribution does not aim to generalization, meaning the methodology and empirical analysis should be merely appropriate for the context of analysis. On the contrary, manuscripts aiming to provide a theoretical contribution and generalization need to meet more stringent methodological standards.

Hypothesis 4: RGs of high rank journals demand the evaluation of a manuscript's methodological validity in stricter terms than RGs of non-high rank journals

3. Data and methods

3.1 Sample

The sample includes journals in the category "Management" of the Web of Science (WOS) (year 2014) - excluding journals with a focus on Psychology. We browsed journals' websites looking for RG, namely texts or evaluation forms with instructions for the reviewers. When missing, an e-mail was sent to the chief editor (one reminder). We found RGs for 46 journals out of 168, which means that at least one in four journals adopts this type of document. Only a few journals adopt the same RG.

Journals in the sample display similar impact factors compared to the overall sample of management journals in the WOS (median 1.14 vs 1.33). The most common publishers are Wiley-Blackwell (14), Emerald (5), Elsevier (4), SAGE (3), AOM- Academy of Management (3). The journals' rank is based on the Academic Journal Guide (AJG) of the Chartered Association of Business Schools (year 2015). The AJG is a ranking of journals widely employed by scholars and institutions in the field of management, thus providing a suitable proxy of the prestige a journal holds among scholars and reviewers in the field. The AJG 2015 ranks 1,396 journals in five levels: 1, 2, 3, 4 and 4*; 39 of the 46 journals in our sample are included in the AJG 2015. Among the 46 journals in the sample, 24 employ an evaluation form to give instructions for reviewers, 20 journals adopt a text, and two journals employ both. Forms are more common among medium-low rank journals, whereas high rank journals' employ texts or both (Table 1).

Table 1 illustrates the journals in the sample and selected characteristics.

[Table 1 here]

3.2 Coding

The coding phase involved two senior researchers with previous experience in coding, and a PhD student.

The text of the RGs was coded in three steps.

First, the three coders individually analysed the RGs in order to identify broad *macro categories* of instructions and information by their distinct function. The respective coding categories were discussed, leading to the identification of nine macro categories of instructions/information, which can be found in the large majority of the RGs. These are described in paragraph 4.1.

Second, the two senior researchers independently i) identified pieces of text providing instructions or information to reviewers – named '*micro categories'* - and ii) assigned them to one of the nine macro categories. Next, the junior researcher controlled for inter-rater agreement, which was 92.0% on the coded text and 91.8% on their macro category.[3] In case of disagreement, the three coders discussed the discrepancy until ultimately agreeing on a final coding. Micro-categories were most always preserved in their original formulation, and only those conveying very similar instructions were merged. For each journal RG a "1" was used to code if the *micro category* of the instruction was present at least once, and "0" if it was not present.

Finally, micro-categories of instruction were coded regarding the function of peer review. namely: i) *gatekeeping*, i.e. instruction to assess the manuscript based on a given aspect; ii) *not gatekeeping*, i.e. instructions to *not* assess the manuscript based on a given aspect; iii) *developmental*, i.e. instructions to contribute improving the manuscript. Gatekeeping instructions were also coded according to whether they pertained issues of: i) *validity* (e.g. accuracy of the methodology), ii) *novelty* (e.g. theoretical contribution) or iii) *communication*, namely whether a manuscript and its elements are properly expressed, (e.g. quality of writing, "clarity" of the research question, theory, methodology).[4]

3.3 Methods

The empirical analysis is organized into three sections.

First, the nine macro-categories of instructions are described.

The second section provides descriptive statistics on the frequency and distribution of macro categories of instructions, as well as correlations between instructions' and the journals' characteristics.

Third, visual representations, correlations, and non-parametric tests on the frequency of different types of instructions between journals' rank layers are used to explore variations in the content of RG and to test the hypotheses.

4. Empirical analysis

4.1 The content of reviewer guidelines

Box 1 describes the nine macro-categories of information and instructions in the selected RG.

Box 1 – Macro- categories of information or instruction

Journal. Information on the journal area of interest, goal and scope. Often, via a web link to the journal's mission statement or editorial statement. For example: "*read the journal's mission statement"*

When to Review. Instruction on when to accept to review or not. For instance: "*a reviewer who feels inadequately qualified to judge the research (...) should refuse to review*"

[3] The coders identified a total of 825 micro-instructions (some repeated more than once in each text); in 66 cases the micro instructions were coded by only one of the coders, inter-rater agreement: (825-66)/825=0.920. It was agreed to retain 774 instructions; in these cases, the coders disagreed 63 times, inter-rater agreement: (774-63)/774=0.918

[4] In some cases the differences between categories are subtle. For instance, assess whether "the knowledge gap is identified" (communication) serves a different purpose from assessing whether the manuscript "fills a knowledge gap" (novelty). Asking to control whether "p-values are presented" (communication) differs from asking whether the "p-values are significant" (validity). Three micro-categories of instruction (5 occurrences) could not be attributed clearly to one of the three categories. For instance, "assess whether the limitations are identified (and discussed)"

Relationships and Roles. Instruction on the competences, duties, and role of reviewer, author and editor and instructions on the appropriate ways to communicate between them. For instance: reviewers should *"avoid personal biases", "respect the confidentiality of the review process',* and *"not provide an editorial opinion (accept, reject, revise) to the author"*

Structure. Instructions on the length and sections of the review, order and number of the comments. For example: "*be 2-4 pages in length"*

Style. Instructions on how to write the review, such as to "*be concise*", "*be clear*", which do not directly impact on the relationship with the author (e.g. "*be polite*")

Gatekeeping. Elements that the reviewer should consider when judging the article's worthiness. For instance: "*does the manuscript test, create, or extend theory?*"

Not Gatekeeping. Elements, which the reviewer should *not* consider for comments or as a criterion for judging whether the article should be accepted. For instance: when evaluating submission from non-native speakers "*distinguish between the quality of writing and the quality of the ideas(…) Those ideas may be good, even if they are not expressed well."(AMJ)*

Developmental. Instructions that the reviewer should contribute to the improvement of the article and/or how they have to do so, such as "*a developmental reviewer (…) determines whether there is a gem of an idea in the manuscript and suggests how the idea might be developed*" (*JIBS*)

Timing. Information/instructions on when to return the review, limits of time, etc. For instance: *"reviewers should be prompt with their reviews"*

4.2 Variations in reviewers' instructions by journals' rank

4.2.1 Descriptive statistics

Table 2 displays the number and frequency of macro and micro categories of instructions, for the overall sample and along the rank levels. On average, the RGs include 16.8 instructions. Gatekeeping instructions are the most diverse and frequent (106 types, 427 occurrences in 45 journals), followed by instructions on the relationships and roles of editors, reviewers and authors (44 types, 138 occurrences in 28 journals), developmental instructions (20 types, 99 occurrences in 30 journals) and non-gatekeeping instructions (8 types, 21 occurrences in 17 journals). RGs of high rank journals (4 and 4*) include a greater number of instructions, particularly regarding developmental instructions and regarding relationships and roles, whereas gatekeeping instructions are less numerous for the very high ranked journals (4*).

[Table 2 here]

Table 3 presents the Pearson correlations between journals' rank level and the frequency of different types of instructions. Journals' rank[5] correlates strongly and significantly with the number of developmental instructions (0.619**), which is consistent with the first hypothesis, and negatively with the number of gatekeeping instructions (-0.358*), and in particular with instructions regarding communication issues (-0.462**), whereas correlations with instructions related to validity and novelty are negative but not significant.

[Table 3 here]

4.2.2 Developmental instructions for reviewers

Developmental instructions often consist of generic invitations "to be developmental", "to help improve" and "give suggestions". Generic developmental instructions are found in 26 RGs. More specific developmental

[5] Level 4* is transformed to 5, whereas journals not in the AJG are not considered.

instructions prescribe to "identify" or "help" type of action, such to "identify strengths/redeeming qualities" (14 RGs), "identify weaknesses/areas for improvement" (13) "help solve problems/weaknesses and point out solutions/offer suggestions" (12) as well as "suggest relevant/useful works/citations/ references" (10). 16 RGs do not include any developmental instructions, whereas 16 RGs include only one or two such instructions, typically generic ones. 14 journals' RGs include four or more developmental instructions; among them, all the seven 4* journals, which display a significantly higher number of developmental instructions compared to the other journals (Mann-Whitney two-tailed p-value 0.0005***).[6]

Figure 1 plots the journals' according to the number of developmental and gatekeeping instructions in the RG, as well as their rank, and clearly shows that very high rank journals (4*) to include a large number of developmental instructions.

[Figure 1 here]

4.2.3 Gatekeeping and not gatekeeping instructions for reviewers

Gatekeeping instructions are included in all of the RGs (with one exception), with an average of 9.63 occurrences. RGs of very high rank journals (4*) display a significantly lower number of gatekeeping instructions (mean 5.3 vs. 10.1, -46%; p-value 0.03), and particularly a smaller number of instructions related to communication issues.

Table 4 presents the twenty most common gatekeeping instructions, which represent 52% of the total gatekeeping instructions, classified by their pertinence to novelty, validity and communication issues, and their frequency for each rank level. Instructions targeting a manuscript's novelty are the most frequent (163), followed by validity (146) and communication (118).

[Table 4 here]

RGs of high rank journals (levels 4 and 4*) instruct to evaluate the theoretical contribution more frequently than non-high rank journals (p-value Mann-Whitney two-tailed 0.08), and more frequently than to evaluate the empirical contribution (for rank 4* journals 43% vs. 14% and for rank 4 journals 80% vs. 60%). The opposite occurs for RGs of non-high rank journals (ranks 1 to 3) in which instructions to assess the practical contribution are comparatively more frequent. These findings are consistent with the second and third hypotheses.

Hypothesis 4 states that RGs of high rank journals instruct reviewers to evaluate a manuscript's methodological validity in stricter terms than non-high rank journals. Instructions related to the evaluation of the methodology employ different wordings. Two instructions refer to methodological 'soundness' and 'rigor', which imply a strict assessment,[7] whereas two instructions employ mild requirements, namely whether the methodology is '*appropriate'*, synonymous being 'suitable or proper in the circumstances, fitting, apt, pertinent to the purpose', and whether the methodology is '*described/explained clearly, adequately'*, namely 'sufficiently, appropriately, suitably'. RGs of high rank journals instruct reviewers to assess whether the methodology is *sound* and/or *rigorous* much more frequently than non-high rank journals (Table 5), thus supporting hypothesis 4.

[Table 5 here]

Eight micro-categories of instructions ask reviewers to refrain from judging a manuscript according to certain criteria, and they occur 21 times. Non-gatekeeping instructions are slightly more frequent among high rank journals (rank 4* mean 0.7 and rank 4 mean 0.8) than in non-high rank journals (rank 3: 0.36; rank 2: 0.50;

[6] Journals not in the AJG are excluded from the non-parametric tests

[7] Synonyms of 'sound' include 'valid, logical, solid, authoritative', and 'rigorous' indicates 'extremely thorough and careful, scrupulously accurate, precise, meticulous, exact, correct, perfectionist, strict'.

rank 1: 0.25). The most common "non-gatekeeping" instructions ask reviewers to not reject an article based on the quality of "*English, writing style, grammar*" (10) and to "*be open-minded about theory, methodological choices*" (4).

It is important to notice that only a handful of instructions relate to some crucial challenges of modern science, related to research ethical behaviours (e.g. Fanelli 2009) and the so-called 'reproducibility crisis' (Baker 2016). Only two instructions regard ethical issues, namely academic misconduct (3 occurrences), and unethical practices (1), while instructions aimed to favour replication only occur in five instances.

5. Conclusions and discussion

This article aimed to address an important gap in peer review research, namely variations in how journals try to shape peer review, by analysing instructions for reviewers included in the reviewer guidelines (RG) of 46 management journals. The analysis shed light on large variety of instructions and evaluation criteria that can define reviewers' goals and on variations among journals of different rank. RGs of very high rank journals include much less gatekeeping instructions and four times more developmental instructions, which are instead barely absent in non-high rank journals RGs. RGs of high rank journals also give more emphasis to theoretical contributions than empirical contributions and instruct to assess methodological validity in stricter terms.

The association between journals rank' and developmental approach in peer review may appear efficient as more time and effort are allocated to review contributions that are potentially more valuable. At the same time, two main issues are worth to be further examined.

First, such association may represent another example of cumulative advantage in science (Merton 1968). In fact, the scientists most likely to undergo the peer review process of top-tier journals and receive more constructive reviews are scientists with the greatest amount of institutional and personal resources. On the contrary, less experienced researchers and from peripheral scientific institutions and systems are more likely to target low rank journals, where reviewers are not expected to be developmental.

Second, the effects of a developmental approach are not known. Future research should explore whether a developmental peer review truly increases the capability to identify and/or recover innovative contributions, at what price e.g. in terms of lengthier review or harming authorial voice (Bedeian 2004), and whether developmental instructions can or should be adopted by non-high rank journals as well.

A related issue is whether reviewers' behaviour can and should be steered (Reale and Zinilli 2017) and how to best balance gatekeeping and developmental functions, focus on validity and novelty (Luukkonen 2012). Kayes (2002), for example, encouraged the adoption of diverse perspectives and blurring theoretical representations in order to detect the most innovative works, while the *Journal of Vocational Behavior* adopts a two-step procedure whereby reviewers first evaluate the research question and methodology, and only in case of "in-principle accept" they consider results and conclusions (Franklin 2017).

A final reflection emerges from contemplating the great variety of evaluation criteria that can guide reviewers behaviour. A renown concern to peer review quality and legitimization is the low level of agreement between reviewers, known as "the luck of the reviewer drawn" (Cole and Simon 1981). Possibly, disagreement between reviewers relates to the implicit adoption of different approaches and criteria. Hence, future research can explore whether instructions for reviewers' may be helpful in addressing "the luck of the reviewer drawn" by fine-tuning what and how reviewers evaluate.

References

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PloS one*, 4(6), e5910.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604), 452.

Balietti, S., Goldstone, R. L., & Helbing, D. (2016) 'Peer review and competition in the Art Exhibition Game', *Proceedings of the National Academy of Sciences* 113/30: 8414-8419.

Bhattacharjee, Y. (2013) 'The mind of a con man', *The New York Times* 28.

Bedeian, A. G. (2004) 'Peer review and the social construction of knowledge in the management discipline', *Academy of Management Learning & Education* 3/2: 198-216.

Bornmann, L. (2008) 'Scientific peer review: An analysis of the peer review process from the perspective of sociology of science theories', *Human Architecture* 6/2: 23.

Campanario, J.M., (1996) 'Have referees rejected some of the most-cited articles of all times?', *Journal of the Association for Information Science and Technology* 47/4: 302-310.

Campanario, J.M., (1998a) 'Peer review for journals as it stands today—Part 1', *Science communication* 19/3: 181-211.

Campanario, J.M., (1998b) 'Peer review for journals as it stands today—Part 2', *Science Communication* 19/4: 277-306.

Campanario, J.M. (2009) 'Rejecting and resisting nobel class discoveries: Accounts by Nobel laureates', *Scientometrics* 81/2: 549–565.

Castellucci, F. and Gokhan, E. (2010) 'What's in it for them? Advantages of higher-status partners in exchange relationships', *Academy of Management Journal,*, 53/1: 149-166.

Chen, J., and Konstan, J. A. (2010) 'Conference paper selectivity and impact', *Communications of the ACM,*, 53/6: 79–83.

Cole, S., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, 214(4523), 881-886.

Corley, K.G. and Gioia, D.A. (2011) 'Building theory about theory building: what constitutes a theoretical contribution?', *Academy of management review* 36/1: 12-32.

Czarniawska, B., and Joerges, B. (1996) 'Travels of ideas'. In: Czarniawska B. and Sevo G. (ed.) *Translating Organizational Change*, pp. 13-47. Walter de Gruyter: Berlin.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama,* 263(10), 1385-1389.

Ellison, G. T. H. and Rosato, M. (2002) 'The impact of editorial guidelines on the classification of race/ethnicity in the British Medical Journal', *Journal of Epidemiology & Community Health* 56/2.

Evangelou, E., Siontis, K. C., Pfeiffer, T., & Ioannidis, J. P. (2012). Perceived information gain from randomized trials correlates with publication in high–impact factor journals. *Journal of clinical epidemiology*, 65(12), 1274-1281.

Eysenck, H. J., and Eysenck, S. B. G. (1992) 'Peer review: advice to referees and contributors', *Personality and Individual Differences* 13/4: 393-399.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one*, 4(5), e5738.

Franklin, J. (2017) 'Results masked review: peer review without publication bias', Available at: https://www.elsevier.com/connect/reviewers-update/results-masked-review-peer-review-without-publication-bias

Hackett, E. J., and Chubin, D. E. (2003). *Peer review for the 21st century: applications to education research*. National Research Council workshop

Hambrick, D. C. (2007) 'The field of management's devotion to theory: Too much of a good thing?', *Academy of Management Journal* 50/6: 1346–1352.

Haug, C.J., (2015) 'Peer-review fraud—hacking the scientific publication process', *New England Journal of Medicine*, 373/25: 2393-2395.

Kalleberg, A.L. (2012). 'Social Forces at 90', *Social Forces* 91/1:1–2.

Kayes, D. C. (2002) 'Experiential learning and its critics: Preserving the role of experience in management learning and education', *Academy of Management Learning & Education* 1/2: 137-149.

Knorr-Cetina, K. (1981) *The manufacture of knowledge: an essay on the constructivist and contextual nature of science*. Oxford, UK: Pergamon Press

Kuhn, T. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, T. (1977) *The Essential Tension*. Chicago: University of Chicago Press.

Lamont, M. (2009). *How Professors Think*. Cambridge, MA: Harvard University Press

Langfeldt, L. (2006) 'The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessment', *Research Evaluation*, 15/1: 31-41.

Latour, B. (1987) *Science in Action*. Cambridge, MA: Harvard University Press.

Lee, C. et al. (2013) 'Bias in peer review.' *Journal of the Association for Information Science and Technology* 64/1: 2-17.

Legge, K. (2001). 'Silver bullet or spent round? Assessing the meaning of the "high commitment management"/performance relationship'. In: Storey, J. (ed.) *Human resource management: A critical text* pp. 21–36. Thomson Learning: London.

Luukkonen, T. (2012) 'Conservatism and risk-taking in peer review: Emerging ERC practices', *Research Evaluation* 21/1: 48-60.

McCook, A. (2006). Is peer review broken? Submissions are up, reviewers are overtaxed, and authors are lodging complaint after complaint about the process at top-tier journals. What's wrong with peer review?. *The scientist*, 20(2), 26-35.

Merton, R.K. (1968) 'The Matthew Effect in Science', *Science* 159/3810: 56-63.

Merton, R.K. (1973) *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago press.

Miner, J. B. (2003) 'The rated importance, scientific validity and practical usefulness of organizational behavior theories: A quantitative review', *Academy of Management Learning and Education* 2/3: 250–268.

Moran, G. (1998) *Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication.* Greenwich, CN: Ablex.

Patriotta, G. (2017) 'Crafting papers for publication: Novelty and convention in academic writing', *Journal of Management Studies* 54/5: 747-759.

Patterson, D.A. (2004) 'The health of research conferences and the dearth of big idea papers', *Communication ACM* 47/12: 23–24

Reale, E. and Zinilli, A. (2017) 'Evaluation for the allocation of university research project funding: Can rules improve peer review?', *Research Evaluation* 26/3: 190-198.

Romanelli, E. (1996). 'Becoming a reviewer: Lessons somewhat painfully learned'. In: Frost P.J. and Taylor M.S. (ed.) *Rhythms of academic life: Personal accounts of careers in academia*, pp. 263–268. Sage: Thousand Oaks, CA.

Sandström, U., & Hällsten, M. (2007). Persistent nepotism in peer-review. *Scientometrics*, 74(2), 175-189.

Sarigöl, E. et al. (2017) 'Quantifying the effect of editor–author relations on manuscript handling times', *Scientometrics* 113/1: 609-631.

Schminke, M. (2002) 'From the editors: Tensions', *Academy of Management Journal* 45/3: 487–490.

Seeber, M. and Bacchelli. A. (2017) 'Does single blind peer review hinder newcomers?' *Scientometrics* 113/1: 567-585.

Siler, K. et al. (2015) 'Measuring the effectiveness of scientific gatekeeping', *Proceedings of the National Academy of Sciences* 112/2: 360–365.

Siler, K. and Strang D. (2017) 'Peer review and scholarly originality: Let 1,000 flowers bloom, but don't step on any', *Science, Technology, & Human Values* 42/1: 29-61.

Smith, R., (2006) 'Peer review: a flawed process at the heart of science and journals', *Journal of the Royal Society of Medicine* 99/4: 178-182.

Squazzoni, F., Brezis, E., & Marusˇic´, A. (2017). Scientometrics of peer review. *Scientometrics*, 113(1), 501–502.

Starbuck, W. H. (2003). 'How much better are the most prestigious journals? The statistics of academic publication', Unpublished manuscript, New York University.

Warren, L. (2003) 'Galileo didn't publish his observations in scholarly journals', *National Geographics* 203/5: 15.

Weller, A. C. (2001) *Editorial peer review: Its strengths and weaknesses.* Medford, New Jersey: Information Today, Inc.

Ziman, J. M. (1984) *An introduction to science studies: The philosophical and social aspects of science and technology*. Cambridge, England: Cambridge University Press.

Zuckerman, H. and Merton, R.K. (1971) 'Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system' *Minerva* 9/1: 66-100.

Zuckerman, H. (1977) *Scientific elite: Nobel laureates in the United States*. London: Transaction Publishers.

Figure 1 - Number of developmental and gatekeeping instructions in the RGs

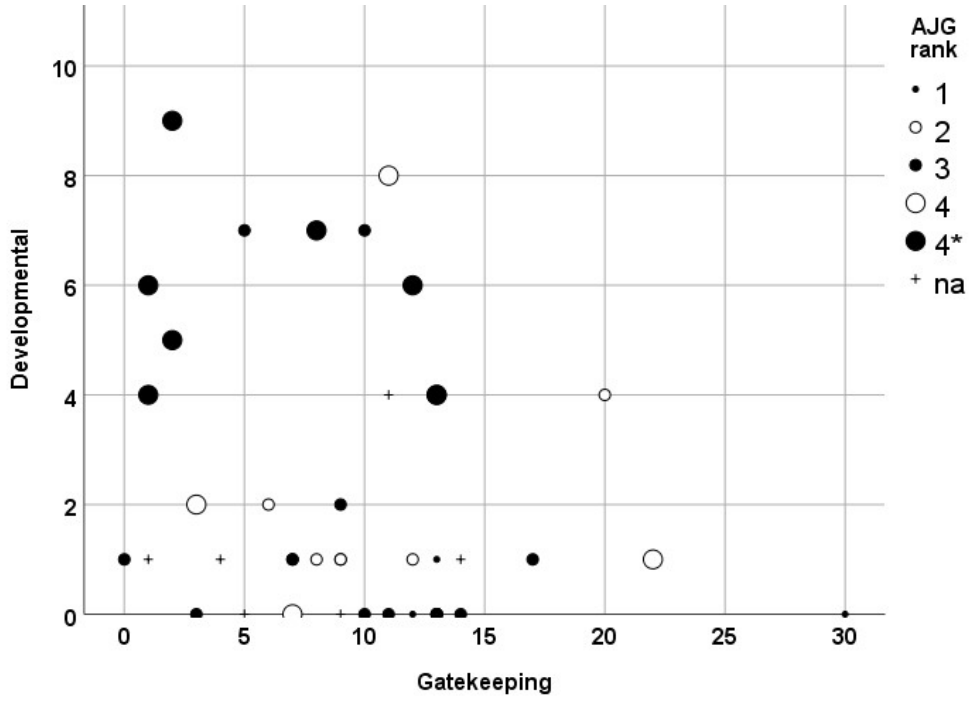

Table 1 - Journals in the sample

Journal	Journal acronym	Publisher*	Rank AJG 2015	Impact Factor*	Reviewer Guidelines
ACAD MAN J	AMJ	Aom	4*	5	text & form
ACAD MANAG LEARN EDU	AMLE	Aom	4	2,1	text
ACAD MAN REV	AMR	Aom	4*	7,8	text
ADMIN SCI QUART	ASQ	Sage	4*	2,4	text
CULT ORGAN	C&O	Taylor & Francis	2	0,5	text
CORP GOV-OXFORD	CG	Wiley	3	1,8	text
INT J CONTEMP HOSP M	CHM	Emerald	3	1,6	form
CAN J ADM SCI	CJAS	Wiley	2	0,6	text
CALIF MANAGE REV	CMR	University of California	3	1,9	text
EMJ-ENG MANAG J	EMJ	Amer Soc Eng Manag	na	0,3	text
EMPL RELAT	ER	Emerald	2	0,4	form
GROUP ORGAN MANAGE	GOM	Sage	3	1,5	text
HUM RELAT	HR	Sage	4	1,9	form
HUM RESOUR DEV Q	HRDQ	Wiley	2	0,9	form
HUM RESOUR MANAG J	HRMJ	Wiley	na	2,4	form
INT J ARTS MANAG	IJAM	Ecole des haute	na	0,2	form
INT J MANAG REV	IJMR	Wiley	3	2,7	form
INT J PROJ MANAG	IJPM	Elsevier	2	1,8	text
IND MARKET MANAG	IMM	Elsevier	3	1,9	form
J BUS LOGIST	JBL	Wiley	2	2,9	form
J CONTING CRISIS MAN	JCCM	Wiley	na	1,1	form
J INT BUS STUD	JIBS	Palgrave	4*	3,6	text
J KNOWL MANAG	JKM	Emerald	2	1,3	form
J MANAGE STUD	JMS	Wiley	4	3,3	text & form
J ORGAN BEHAV MANAG	JOBM	Taylor & Francis	2	0,4	form
J ORGAN END USER COM	JOEUC	Iigi Publ	1	0,4	form
J OPER MANAG	JOM	Elsevier	4*	4,5	text
J OPER RES SOC	JORS	Palgrave	3	0,9	form
J PROD INNOVAT MANAG	JPIM	Wiley	4	1,4	text
J SUPPLY CHAIN MANAG	JSCM	Wiley	3	3,7	text
J SERV MANAGE	JSM	Emerald	2	1,3	form
J SPORT MANAGE	JSM	Human Kinetiks	2	0,7	form
MANAGE INT REV	MIR	Springer	3	0,9	form
MANAGE ORGAN REV	MOR	Wiley	3	3,3	text
MANAG SERV QUAL	MSQ	Emerald	1	1	form
NONPROFIT MANAG LEAD	NPM	Wiley	1	0,4	text
ORGAN BEHAV HUM DEC	OBHDP	Elsevier	4	2,9	text
OPER RES	OR	Informs	4*	1,5	text
RAE-REV ADMIN EMPRES	RAE	Fundacao getulio	na	0,2	text
RBGN-REV BRAS GEST N	RBGN	Fund escola comercio	na	0	form
R&D MANAGE	RDM	Wiley	3	1,3	form
RES TECHNOL MANAGE	RTM	Industrial research	2	0,8	form
S AFR J BUS MANAG	SAJBM	Assoc Prof Man	1	0,2	form

STRATEGIC MANAGE J	SMJ	Wiley	4*	3	text
SCI PUBL POLICY	SPP	Ooxford University press	2	1	form
Z PERSONALFORSCH	ZPF	Rainer hampp verlag	na	0,8	form

* year 2014

Table 2 - Descriptive statistics of macro-categories and micro-categories of instructions

macro category	types of micro categories	total number of occurrences	RG with this kind of instruction	sample mean	4*	4	3	2	1	na
Journal	1	15	33%	0,3	0,4	0,8	0,3	0,3	0	0,1
When	1	13	28%	0,3	0,7	0,4	0,3	0,2	0	0,1
Relationships and Roles	44	138	61%	3	7	3,8	2,8	1,8	1,3	1,7
Structure	1	18	39%	0,4	0,9	0,6	0,5	0,2	0,3	0,1
Style	5	19	30%	0,4	0,7	1	0,2	0,4	0,3	0,1
Gatekeeping	106	427	98%	9,3	5,3	11,2	9	9,9	15,8	7,7
Not Gatekeeping	8	21	37%	0,5	0,7	0,8	0,4	0,5	0,3	0,1
Developmental	20	99	65%	2,2	5,9	3	1,7	1,3	0,3	1
Time	2	19	41%	0,4	0,7	0,6	0,5	0,3	0,3	0,1
Total	188	769		16,8	22,3	22,2	15,7	14,9	18,6	11

Notes: for Journal, When and Structure we did not distinguish different micro-categories of instructions

Notes: for Time we distinguish between time specific (e.g. within 2 months) and generic (e.g. in prompt time)

Table 3 - Pearson correlations between selected journals' characteristics and frequency of instructions in the RG (n. = 39)

	AJG rank	Total instructions	Developmental	Gatekeeping	Not Gatekeeping	Validity	Novelty	Communication
AJG rank	1	0,299	,619**	-,358*	0,198	-0,297	-0,127	-,462**
Total instructions	0,299	1	,638**	,474**	,374*	,482**	,450**	0,25
Developmental	,619**	,638**	1	-0,251	,400**	-0,162	-0,174	-,316*
Gatekeeping	-,358*	,474**	-0,251	1	-0,253	,940**	,769**	,801**
Not Gatekeeping	0,198	,374*	,400**	-0,253	1	-0,259	-0,173	-0,203
Validity	-0,297	,482**	-0,162	,940**	-0,259	1	,663**	,692**
Novelty	-0,127	,450**	-0,174	,769**	-0,173	,663**	1	,304*
Communication	-,462**	0,25	-,316*	,801**	-0,203	,692**	,304*	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 4 – Most common gatekeeping instructions by status layer
N=Novelty, V=Validity, C= Communication

		sample		4*	4	3	2	1	na
total gatekeeping		**9,3**		5,3	11,2	9	9,9	15,8	7,7
validity (mean)		**3,2**		1,7	4,2	3,2	3	6	2,6
novelty (mean)		**3,5**		2,6	4,6	3,7	3,6	4,3	3
communication (mean)		**2,6**		1	2,4	2,1	3,3	5,5	2,1
1	empirical contribution (N)	23	50%	14%	60%	55%	50%	75%	57%
2	coverage of relevant literature (V)	22	48%	14%	20%	45%	58%	100%	57%
3	fit to journal (C)	17	37%	29%	20%	45%	33%	50%	43%
4	quality of writing (C)	17	37%	0%	40%	18%	58%	75%	43%
5	theoretical contribution (N)	16	35%	43%	80%	36%	25%	25%	14%
6	originality (N)	14	30%	0%	20%	36%	58%	25%	14%
7	generic (assess the contribution/significance) (N)	13	28%	14%	40%	27%	17%	75%	29%
8	appropriate methodology / design (V)	12	26%	0%	60%	18%	42%	25%	14%
9	literature accurate/adequate understanding (V)	10	22%	14%	20%	18%	33%	50%	0%
10	evidence based conclusions (V)	9	20%	0%	0%	36%	8%	50%	29%
11	implication for research (N)	9	20%	29%	0%	27%	25%	25%	0%
12	advance knowledge (N)	8	17%	0%	40%	9%	8%	25%	43%
13	quality of theoretical framework (V)	7	15%	14%	40%	9%	0%	25%	29%
14	results analysed appropriately (V)	7	15%	0%	0%	18%	25%	50%	0%
15	interesting (N)	7	15%	29%	20%	9%	17%	0%	14%
16	topic important and interesting (N)	7	15%	29%	0%	9%	17%	25%	14%
17	integration/synthesis existing theories/research (N)	7	15%	14%	40%	18%	8%	25%	0%
18	structure, organization of the manuscript (C)	7	15%	0%	20%	18%	8%	50%	14%
19	methodology explained/clear/adequate descript. (C)	7	15%	0%	20%	9%	25%	50%	0%
20	clarity of the results (C)	7	15%	0%	0%	9%	42%	25%	0%

Table 5 – Instructions related to the assessment of the methodology

Instruction	Whole Sample	4*	4	3	2	1	na
appropriate methodology (V)	26%	0%	60%	18%	42%	25%	14%
methodology explained/ clear/adequate description (C)	15%	0%	20%	9%	25%	50%	0%
methodological soundness (V)	13%	29%	40%	9%	8%	0%	0%
methodological rigor (V)	13%	14%	20%	18%	0%	0%	29%
N. methodological instructions (mean)	0,67	0,43	1,40	0,54	0,75	0,75	0,43