

科学计量视角下的论文同行评议研究综述^{*}

■ 张光耀^{1,2} 谢维熙¹ 姜春林¹ 王贤文¹

¹ 大连理工大学科学与科技管理研究所暨 WISE 实验室 大连 116024

² 荷兰马斯特里赫特科技创新经济研究院(UNU-MERIT) 马斯特里赫特 6211AX

摘 要: [目的/意义]旨在通过梳理科学计量视角下同行评议相关研究,为深入开展同行评议研究提供参考。[方法/过程]在概括同行评议量化特征的基础上,从同行评议公平性、可靠性、预见性 3 个方面对国内外研究现状进行综述。[结果/结论]总结科学计量视角下的论文同行评议研究在研究方法多样性和数据应用全面性方面的趋势,并指出当下研究在结论的可解释性上仍然有待提升、国内同行评议实证研究相对不足等问题。

关键词: 同行评议 开放同行评议 科学学 科学计量学

分类号: G250

DOI: 10.13266/j.issn.0252-3116.2022.14.014

1 引言

同行评议是科学的基石,在科学交流体系中发挥着重要作用。一方面,同行评议作为质量控制的主要机制,可以协助编辑做出论文发表决定^[1];另一方面审稿人在评议过程中针对文稿提出意见,有助于提升论文质量^[2]。在近年颁布实施了一系列政策文件,如中共中央办公厅和国务院办公厅《关于深化项目评审、人才评价、机构评估改革的意见》、科技部《关于破除科技评价中“唯论文”不良导向的若干措施(试行)》、教育部和科技部《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》的大背景下,同行评议的重要性越发凸显。尽管同行评议在科学发展中的地位如此重要,然而学术界对其认识依然有限,因此,同行评议也成为情报学、科技政策、出版学研究者的重要关注方向^[3]。

相比国际上不断增多的同行评议研究^[4],国内同行评议研究更倾向于经验性的阐释,正如 M. S. Dalton 对国际上早期同行评议研究的评价,“大多数关于期刊同行评议的论文往往是编辑的经验性思考,更多的是观点而不是研究”^[5]。虽然已经有学者认识到了同行评议数据,包括送审量和审稿周期等信息的价值,并展开了统计分析,但相比国际上涌现的大量实证研究仍显得不足。此外,国内关于同行评议研究的综述梳理,

更大程度上从出版学的视角出发,其学术贡献更多地体现在介绍与总结国外同行评议经验方面^[6-8]。尽管已有学者注意到同行评议模式的研究变化趋势以及同行评议模式的实际应用情况^[9],但国内学界对科学计量视角下的论文同行评议实证研究仍然关注较少。

从现有文献对同行评议的探讨来看, L. Bornmann 曾提出研究同行评议的 3 个关键点^[1]:①公平性,即某些作者群体在同行评议中是否被偏爱或者受歧视?(Are certain groups of applicants or authors favored or at a disadvantage?);②可靠性,即同行评议的结果是可靠的还是纯属偶然?(Is the selection of scientific contributions reliable or is the result purely incidental?);③预见性,即同行评议结果是否与之后的文献计量表现有关?(Do the selection decisions correlate with scientific performance measures subsequent to decision?).国际上也有学者对同行评议予以高度重视,并开展了大量与之相关的实证研究^[4],其成果也多发表在 *Scientometrics*、*Journal of Informetrics* 等国际科学计量主流期刊上^[10]。此外,基于同行评议数据开展的实证研究为同行评议的科学化提供了重要证据^[11]。从实证研究中获得的发现可以反哺同行评议建设,对于提升同行评议的效率和可信性有着重要价值^[7]。因此,本研究在同行评议量化特征揭示的基础上,依托 L. Bornmann 提出的框架,从同行评议公平性、可靠性、预见性 3 个方面展

^{*} 本文系国家自然科学基金项目“科学文献全景大数据下的研究热点及研究前沿探测”(项目编号:71974029)研究成果之一。

作者简介: 张光耀,博士研究生;谢维熙,博士研究生;姜春林,教授;王贤文,教授,博士,通信作者,E-mail:xianwenwang@dlut.edu.cn。

收稿日期:2022-02-17 修回日期:2022-05-10 本文起止页码:137-149 本文责任编辑:易飞

开论文同行评议研究综述,力图梳理该领域的研究进展,为后续的相关研究提供参考和借鉴。

考虑到研究的话题性与专业性,本研究采取如下检索策略和人工识别方法获取论文同行评议研究相关文献:2021 年 10 月-12 月,以“TS = peer review”为检索式在 Web of Science 中检索,将文献来源限定为知名综合性学术期刊 *Nature*、*Science*、*PNAS*,获得文献 278 篇;将文献来源限定为科学计量学三大主流期刊 *Scientometrics*、*Journal of Informetrics*、*Journal of the Association for Information Science and Technology*,获得文献 390 篇,两部分文献经筛选后获得 118 篇文献。以获取的文献作为种子文献,在研读文献的过程中,根据引用与被引用文献进行一级引文扩展检索,进一步扩展文献范围。在中文论文获取方面,在 CNKI 中将来源类别限定为 CSSCI 期刊,并以“同行评议”“同行评审”“审稿”为主题词进行主题检索,获得文献 4 405 篇,对检索得到的文献进行人工识别后进一步筛选出 121 篇文献。

2 同行评议量化特征揭示

同行评议量化特征是同行评议计量研究的基础,研究者梳理和总结了同行评议量化特征,并在此基础上进一步将研究重点拓展到学术界所关注的公平性、可靠性和预见性。然而国内对该类研究仍然较少涉及,因此,对于量化特征的总结有助于学术界了解该领域的研究现状。此外,鉴于数据的可获取性,评审报告和审稿人往往是研究者最为关注的部分。因此,本部分将对评审报告和审稿人相关研究进行梳理。

2.1 评审报告分析

2.1.1 审稿意见长度

在揭示论文同行评议量化特征的研究中,审稿意见长度和审稿速度是最容易切入的研究点。D. N. Laband 在 1990 年提出将审稿意见长度与质量联系起来^[12],而后续的研究明显接受了这一观点,并提出审稿意见长度反映了审稿意见中所包含信息量的多少^[13-14]。他们认为,一份详细的报告不仅应当包含更多有价值的信息,帮助编辑判断稿件的质量、严谨性和有效性,还可以为作者提供改进研究的建议。N. Casnici 等在分析 *Journal of Artificial Societies and Social Simulation* (JASSS) 的审稿意见时,认为长度可以反映审稿意见质量^[13]。此外,F. Bianchi 等在研究中将审稿意见长度作为评价审稿人的维度之一^[14]。

当考虑多个维度时,审稿意见长度则体现出更丰富的特征^[15]。G. Zhang 等分析了 Publons 平台审稿意

见的长度分布,并讨论了其与性别、国家经济背景和文化间的关系^[16],指出来自具有更完善同行评议系统的发达国家的审稿人可能经历更规范的审稿训练,倾向于写出更长的审稿意见。I. Buljan 等分析 PEERE 数据库的 472 449 意见后发现男性写的意见更长^[17],而 L. Zhang 等在分析 Publons 平台“top reviewer”数据后得出了与前者相反的结论^[18]。从年龄上来看,年轻审稿人可能写出更长^[19]、更有建设性的意见^[20]。比较审稿意见长度也被用于探讨同行评议模式和学科间差异,如 L. Bornmann 等发现开放同行评议下的审稿意见比封闭同行评议下的意见更长^[21],G. Zhang 等和 I. Buljan 等的研究都揭示了社会科学期刊有着更长的审稿意见^[16-17]。

2.1.2 审稿速度

科研评价体系不仅要求作者发表更多论文,也要求作者尽快地发表论文。审稿速度不仅关系到出版延迟^[22],也成为作者投稿时的重要考量因素^[23]。因此,审稿速度是研究者评价同行评议质量的指标之一^[13-14]。B. Bilalli 等在 2020 年开发了自动提取论文审稿速度的系统,此举也在一定程度上表现了学术界对审稿速度的关切^[24]。

随着相关研究的深入,与审稿速度相关的影响因素不断得到揭示。从学者身份和社会关系来看,学者在自己担任编委的期刊发表论文,可以更快被发表^[25],作者和编辑间合作关系能加速论文的处理^[26];与编辑建立人际关系的审稿人会更快地完成审稿工作^[27]。从审稿人性别来看,M. Davo 等针对 *Journal of Epidemiology and Community Health* 的分析发现,女性审稿人提交审稿意见的速度比男性更快^[28],然而,这一结论在对更大样本的分析中出现了差异,K. B. Schmaling 等对 2 178 名审稿人开展的调查显示,女性完成审稿的时间(28.9 天)明显长于男性(26.7 天)^[29]。

审稿速度也存在学科差异,具体而言,自然科学相较于社会科学审稿更快^[30],这可能是由于社会科学有着更加严格的审稿标准,而且从事社会科学研究的学者更倾向于撰写较长的审稿意见^[17]。从期刊层面而言,对于最终录用的论文,影响因子高的期刊审稿速度更快,这可能因为高影响因子期刊有更健全的同行评议系统,可以迅速拒绝学术贡献较低的来稿,而且更容易找到审稿人^[31]。K. Okike 等发现相比于低影响因子期刊(27.6 个月)和专业期刊(21.8 个月),高影响因子的综合性期刊有着更快的同行评议速度(19.1 个

月)^[32]。从论文自身而言,“发表偏倚”的存在使期刊更喜欢发表阳性结果的论文,而且发表速度更快^[33]。

审稿速度也受学术环境和社会环境变化的影响, D. A. Charen 等分析 2010-2015 年间的骨外科手术文献后发现审稿速度出现了加快趋势,这可能得益于编辑们的重视和审稿人熟练度的提升^[34]。S. P. J. M. Horbach 分析了 14 种医学期刊中的 669 篇文章后发现疫情背景也大大加速了 COVID-19 相关研究的发表速度,然而这也引发了学界对同行评议在论文质量控制上的担忧^[35]。

值得注意的是,目前对于审稿速度的研究往往根据论文提交与修回时间,或提交与出版时间测度。然而,即使是修回时间,期刊往往仅提供最终修回时间,这可能会在一定程度上影响研究的准确性。

2.1.3 审稿意见文本

审稿意见长度和审稿速度仅反映同行评议的基本特征,作为一种特殊的语料库,同行评议报告包含了丰富的评审信息,引起了包括计算机和图情研究者的广泛关注。在开放评审报告的大背景下,随着自然语言处理和情感分析技术越发普及,研究者在打开同行评议过程黑箱方面进行了大量研究。

在审稿意见的语言特质分析方面, A. F. Delgado 等基于 *BMJ* 提供的评审报告发现英文审稿意见文本在词汇的使用层面具有显著的量化特征,审稿人更关心方法,而非结果^[36], G. Zhang 等则在对 *BMJ* 评审报告的分析中进一步揭示了“however”这一特殊评审符号的特征和功能^[37]。A. Matsui 等分析 *PeerJ* 的评审报告后发现,在评审报告中署名的审稿人会写出更加积极的意见^[2]。张明阳等重点关注了 *Openreview* 平台中国审稿人的语言风格和打分情况,发现中国学者可能对作者更加宽容^[38]。利用大规模数据集, I. Buljan 等发现审稿人推荐对评审报告的语言特征(和长度)影响最大^[17]。

特定事件背景下的审稿意见也揭示了特殊的评审要求, S. Horbach 等分析 *BMJ* 的审稿意见后发现审稿人在疫情期间对论文的要求放低^[39], 相比于非 Covid-19 研究, Covid-19 研究更少被要求补充额外的实验和进行重大修改, 简单承认研究局限或重新表述结论即可被审稿人接受。

在揭示同行评议报告结构方面, 秦成磊等曾在研究中重点关注了文本挖掘在同行评议中的应用^[9], 例如分析论点挖掘框架下同行评议报告的内容和结构^[40], 探讨审稿意见在学术论文不同结构功能中的分

布^[41]。M. Meng 等提出了一种构建学术文章多层次互评的方法, 为研究者提供了优化论文的方向^[42]。

在分析包括情绪在内的文本特征, 用于论文质量评估和录用结果预测方面, D. Kang 等在公开 *PeerRead* 数据集的同时提出了根据文本特征以及评审分数来预测论文的接收率的构想^[43]。在考虑互联网定性评价信息的基础上, 林原等提出了融合 Twitter 文本情感特征和同行评议情感特征的论文质量评价模型^[44]。从知识发现的角度来看, 研究者观察到审稿意见文本内容与论文内容间存在相关性, 认为同行评议报告可以作为论文内容的补充^[45]。P. Vincent-Lamarre 等发现论文可读性可能与论文录用有关, 可读性较差、使用较多术语的手稿更容易被接受^[46]。

除审稿人撰写的意见文本数据外, 对作者反馈和后同行评议平台评论信息的分析也有助理解同行评议。S. Jiang 从作者的角度出发, 分析作者在 scirev.org 分享的经验, 分析了作者对同行评议过程, 包括决定、时间、评审数量和评审质量的情绪反应^[47]。就后同行评议的研究而言, F. Bordignon 讨论了 PubPeer 在科学中的纠错功能, 比如更正 (corrected) 和撤销 (retracted) 的论文在 PubPeer 上得到的评论比其他论文更多^[48]。J. Ortega 统计了 PubPeer 上的 3 种编辑通告, 勘误 (errata)、关注表达 (expressions of concern) 和撤销 (retractions) 在不同学科下的差异, 发现生物化学、医学和综合期刊中, 编辑通知更为频繁^[49]。J. Ortega 在之后的研究中又进一步分析了 PubPeer 平台的评论类型 (延迟、发帖数量) 的特点和学科差异, 其结果包括超过三分之二的评论讨论了科学中不当行为, 健康科学和生命科学是被讨论最多的领域^[50]。

然而, 值得警惕的是, 机器标注虽然可以支持大样本分析, 但标注准确性与人工标注存在差距^[51]。J. Luo 等指出, 相比于人工编码, 基于词典的情感分析算法 (如 TextBlob 和 VADER) 对同行评议报告的分析难以保证准确性^[52]。

2.2 审稿人分析

2.2.1 审稿人特征揭示

在开放同行评议尚未得到重点关注之前, 期刊编委会这一特殊审稿人群体因其地位的重要性与身份的公开性成为研究者的重点关注对象, 如姜春林等在 2014 年关注期刊的“联锁编委” (interlocking editorship) 现象^[53]。学术界在近几年的研究中也保持了对编委的关注, 包括编委的发文和引用情况^[54-55], 以及编委特征与文献计量指标的关联, 如编委数量与科研

产出评价指标存在正相关关系^[56]、构建期刊编委指数用于机构评价^[57]、编委会的学术表现可以反映期刊声誉^[58]。

随着开放同行评议的发展,审稿人从幕后走向台前,其学术贡献开始受到越来越多的重视^[59],开放同行评议背景下的审稿人研究逐渐成为主流。Publons 平台为在较大数据尺度上探讨审稿人贡献和开展审稿人评价提供了重要数据来源。J. Ortega 曾指出文献计量指标与同行评议活动之间的相关性较弱,说明审稿人可能是独立于研究工作的学术活动^[60]。刘丽萍等发现 Publons 平台优秀审稿人在领域上主要集中于医学和工程等专业,从地理上主要集中于欧美等发达国家^[61]。雷燕证实了医学领域期刊学术影响力与审稿人学术水平及审稿表现呈正相关^[62]。C. Wei 等进一步利用格兰杰检验讨论了审稿人发文与审稿之间的方向性,发现审稿较少或发文较少的科学家在论文评审和发表活动之间存在格兰杰因果关系^[63]。

结合审稿人身份特征分析可以进一步揭示审稿人在同行评议中的差异。在审稿人的学术地位方面,年轻的审稿人提供了更积极、更客观的建议^[19],高级研究人员撰写的意见比初级研究人员更严厉^[13]。在性别方面,G. Bravo 等分析了来自 5 本开放同行评议期刊的 18 525 份评审意见后,发现男审稿人倾向于写更具建设性的报告^[19]。然而,D. Garcia-Costa 等基于更大样本(来自 740 本期刊的 130 万份评审报告)的研究发现,女性审稿人的意见更有建设性^[20]。F. Squazzoni 等发现除物理科学(physical science)外,女性往往比男性提供更加积极的建议^[64]。

2.2.2 审稿人推荐、分配与评价

在科学文献大量增长而审稿人增长有限的情况下,“审稿人危机”成为期刊和出版商高度重视的问题^[65]。一方面,期刊和出版商面临审稿人的补偿^[66]和奖励^[67]难题;另一方面,期刊在挑选审稿人方面存在评审质量和效率上的权衡,招募更多审稿人可以减小评审偏差,但也容易浪费审稿人资源^[68]。因此,如何促进审稿人的推荐与分配也是学界关注的重要话题。Y. Kalmukov^[69]和 G. Chughtai 等^[70]分别从不同角度提出了推荐和分配审稿人的方法。此外,审稿人评价也是期刊重点关注的话题,J. Garcia 等指出稿件质量与期刊标准之间的匹配性可以成为衡量审稿人表现的一个重要指标^[71]。F. Bianchi 等则根据同行评议长度、同行评议持续时间和编辑与审稿人的一致性设计了审稿人评价指标 F3^[14]。

2.2.3 基于 ABM 的审稿人互动模拟仿真分析

鉴于同行评议研究的多样性、复杂性、难以获取数据以及实验成本高等因素^[72],利用计算机模拟来分析和解释社会现象的建模仿真(simulation modeling)也成为同行评议研究领域的重要研究手段。T. Feliciani 等将相关研究划分为评审决定的形成规则、稿件和审稿人的匹配规则、编辑政策、审稿人行为、评审系统的比较和偏见的识别与处理 6 种,其中 Agent-Based Modeling (ABM) 是同行评议仿真研究中使用最多的方法^[72]。与基于实证、大数据的关联性(dependence)视角的观察范式不同,ABM 范式基于生成性(generative)视角观察数据的生产过程,通过纵向视角揭示微观个体互动涌现出的宏观社会过程和社会现象^[73]。

F. Squazzoni 等提出从社会互动的视角来研究同行评议行为,并开展了大量模拟仿真研究,他们的研究揭示了审稿人行为的重要性,包括分析同行评议的运行过程,如由于审稿人的不包容而产生的对创新的抑制^[74];理解科学家的行为动机和后果,增强同行评议的质量、效率和公平性,如 S. Righi 等基于 ABM 关注了审稿人的动机,面对审稿任务与论文写作对精力的竞争,做出低质量的评审成为审稿人的更优选择^[75],F. Squazzoni 等的研究表明只有作者认为自己在同行评议中得到公正对待,在之后担任审稿人时,他才会去公平对待别人^[76];审稿人行为会对同行评议的质量和效率产生巨大的影响,可靠的审稿人和同质性的科学共同体可以让同行评议更加公平,而审稿人的可靠性越高会加剧学术界的马太效应^[77]。J. Bautista 等基于 ABM 重点关注了同行评议中的作弊者,模拟了审稿人在同行评议过程中对于竞争者稿件的压制,这种有意的差评比普通的误判更危及同行评议的质量和效率^[78]。

3 同行评议的公平性

3.1 性别

同行评议的公平性关注某些作者群体在同行评议中是否被偏爱或者受歧视^[1],性别不平等是其中重要的关注点^[79]。性别不平等一方面体现在相比女性在科学中的贡献,女性审稿人在同行评议中的代表性仍显不足^[18]。比如医学期刊中的女性主编占比偏低(444 位医学期刊主编中仅有 94 位是女性,占比 21%)^[80];女性比男性更不愿意接受期刊编委会的邀请。在生态学和进化生物学领域,女性担任审稿人和编辑的比例相比于她们在研究人员中的比例而言仍然

偏低^[81];编辑在选择审稿人时倾向于选择同性别的审稿人^[82]。F. Squazzoni 等发现疫情不仅影响了女性的科研产出,在此期间女性对同行评议邀请的接受率也相对更低,意味着女性更容易受到生活和家庭的影响^[83]。

性别不平等的另一方面体现在女性作者在同行评议中可能受到了不公正的对待。D. Murray 等基于 ELife 的同行评议数据,讨论了作者与审稿人性别同质性与同行评议结果的相关性,具体而言,末位作者为男性的稿件接受率明显高于末位作者为女性的稿件^[84]。B. Demarest 等对 *Journal of the American Society for Information Science and Technology* 评审数据的研究发现女性作者的论文收到的负面评价更多^[85]。C. Fox 等对于 6 本生态学期刊超过 23 000 篇论文的同行评议结果分析也表明,女性作者的论文接受率更低^[86]。F. Squazzoni 等对来自 145 本期刊的 76 万篇同行评议文本,从审稿人的选择、审稿意见和编辑决定 3 个角度分析,发现女性作者的手稿并未在同行评议过程中受到歧视^[64]。然而,D. Card 等分析了提交给 4 本重要经济学期刊的评审数据后发现女性作者的论文被引用的次数比男性作者的论文高 25%,这可能意味着女性作者被施加了更高的标准^[87]。

3.2 国家

从国家层面来看,B. Demarest 等曾指出非西方的学者在英文场域下更有可能得到负面的评论^[86]。C. Fox 等的研究揭示出编辑更喜欢邀请来自北美地区的审稿人^[88];作者的地理位置与评审结果有关,具体而言,来自美国和英国的第一作者的论文得到的评价更好,来自非洲、亚洲和拉丁美洲的第一作者的论文得到的评价更差^[89]。D. Murray 等也指出,审稿人可能偏爱来自本国作者的稿件,作者遇到有相同国籍的审稿人时,论文更可能被接受^[84]。H. Zhu 发现相比于非美国作者,美国作者发表在 *New England Journal of Medicine (NEJM)* 的论文影响显著更低,并讨论了同行评议中可能存在的本国偏好(home country bias)^[90]。X. Zhang 以英文环境学期刊的中国作者的 157 篇稿件为例,以检测在单盲制度下的地理差异,但并没有发现差异^[91]。A. Campos-Arceiz 等的研究发现,相比非中国审稿人,中国审稿人对中国作者更严格^[92]。

3.3 机构

在机构层面,Y. Wei 等比较了哈佛和非哈佛大学在 *NEJM* 发表论文的被引差异,并未发现 *NEJM* 存在机构偏好^[93]。同样关注被引差异,郭峰等分析了国内

经济学期刊中的机构偏好,发现机构偏好导致论文被引下降,但近年来该现象有所缓解^[94]。关于单双盲审稿对比的实验研究,R. Blank 的实验研究表明,相比于单盲审稿,双盲审稿下的论文接受率更低,作者所在的机构的声望和质量可能会带来同行评议中的偏差^[95]。A. Tomkins 等设计了计算机会议论文中的单双盲审稿对比试验,结果发现单盲审稿人比双盲审稿人更倾向于推荐来自知名学者、顶尖大学和顶尖企业的论文^[96]。

3.4 人际关系

人际关系可能会影响同行评议中的公平性。M. Seeber 等基于对计算机会议作者的分析,发现单盲审稿制度可能会阻碍学术新人的发展^[97]。C. Fox 等基于 *Functional Ecology* 期刊的审稿人信息发现作者推荐审稿人会导致编辑决策偏差,具体而言,作者推荐的审稿人比编辑选择的审稿人对论文的评价更高,而且这种推荐增加了论文被接受的可能性^[98]。J. Shopovski 等基于 *European Scientific Journal* 的数据得出了类似的结论,并建议编辑应避免仅靠作者的推荐选择审稿人^[99]。网络分析进一步量化了同行评议中审稿人的“偏爱”,E. Sarigoel 等分析了 2007 年至 2015 年期间发表在 *PLOS ONE* 的 10 万多篇文章,发现作者和编辑间合作关系能加速论文的处理^[26]。M. Teplitskiy 等利用 *PLOS ONE* 的评审数据构建了审稿人-作者网络,发现审稿人倾向于对网络中离自己更近的学者给出更高的分数^[100]。G. Bravo 等也构建了作者-论文网络和审稿人论文网络,根据作者在网络中的地位来测度声望,在论文受到负面评价的情况下,声望较高的作者被编辑拒稿的可能性更小^[101]。

4 同行评议的可靠性

同行评议的可靠性是学术界关注的重要议题^[102]。可靠性关注同行评议的结果是可靠的还是纯属偶然^[1]? D. Cicchetti 曾将审稿人的可靠性定义为“两个或多个审稿人对同一论文的判断达成一致的程
度”^[103],并认为当审稿人之间达成高度一致时,同行评议被认为是可靠的。然而,相关的研究证实评审结果并不可靠,L. Bornmann 等针对审稿人信度(inter-rater reliability)进行的 meta 分析也表明了审稿人之间的低一致性^[104]。在经验研究中,审稿人意见完全相同的期刊占比较低,甚至有 10% 的论文的专家审稿意见完全相反^[105],来自国际放射学、神经科学和康复学的证据也表明论文评审意见不一致的现象时常发

生^[106-107];刘欣等基于 ICLR 会议的公开同行评议数据开展的研究也发现同行评议对于质量较差论文的一致性高于质量较好的论文^[108]。岳名亮等提出了测度审稿人一致性的方法,以提高审稿人遴选效率^[109-110]。

R. Kravitz 等利用 *Journal of General Internal Medicine* 的数据证实了编辑的决定很大程度上受到审稿人意见的影响^[111]。编辑在控制同行评议质量、提升可靠性的过程中发挥着更加关键的作用^[112],特别是在处理审稿不充分、审稿人不可靠和不道德评审的情况下^[113]。正因如此,期刊的异常编辑活动尤其受到学者关注^[114]。

实验研究在探讨同行评议可靠性方面最容易得到出人意料的结果。同行评议研究中知名的“重复投稿实验”,即 D. Peters 等从 12 本心理学知名期刊上挑选了 12 篇著名机构发表的论文重新投稿,结果有 8 篇论文因方法缺陷被拒稿,这引起了学术界对同行评议可靠性的反思^[115]。尽管该实验在当时引起很大的争议,在之后也因抄袭检测系统的发展和科研伦理上的批评而难以复现^[115]。NIPS (Neural Information Processing Systems, 神经信息处理系统大会)在 2014 年将 166 篇论文分别提交两组委员会评审,结果被录用的 37 篇论文中,只有 16 篇论文(占比 43%)同时被两组委员会录用,这一结果表明了同行评议的可靠性低,改变审稿人会对评审结果有很大的影响^[116]。E. Brezis 等模拟了 NIPS 实验,进一步探讨了同行评议中随意性的原因,比如审稿人对创新的认知以及审稿人分配给评审的时间存在差异^[117]。C. Cortes 等在 2021 年重新评估了 2014 年 NIPS 的论文审稿情况,发现以引用量为衡量标准,被接收论文的评分和影响力之间没有相关性,而在被拒收的论文之间存在相关性,说明同行评议更容易识别质量较差的论文,但判断高质量论文的能力一般^[118]。

同行评议的复杂性使得在考虑同行评议可靠性时还需要考虑审稿人之外的因素,如期刊层面的差异,M. Seeber 发现不同级别的期刊对审稿标准和重点的要求不同,低层次期刊要求审稿人以表格形式评估稿件,而高层次期刊的审稿人以文本形式评审稿件,而且对方法有效性的要求更高^[119]。此外,作者对是否在同行评议期刊发表论文也存在选择性,如 D. Sakai 发现日本的年轻学者更喜欢发表同行评议论文,随着学者年龄增长,他们发表同行评议论文的比例也在下降^[120]。

在提升同行评议可靠性方面,S. Horbach 等调查

了不同评审方式下论文的撤稿率(retraction rate)差异^[121],证实盲审制度的实施和检测工具的应用可以降低撤稿率。J. Leek 等将同行评议视作合作过程(cooperative process),实施了作者和审稿人在开放和封闭同行评议中合作的在线博弈实验。结果表明,当审稿人行为被公开时,随着审稿人得到的奖励增加,合作也显著增加,合作的增加可以提升评审的正确性^[122]。对于学术界高度重视的预注册制度(pre-registration),C. Soderberg 等组织的实验研究验证了同行评议过程中预注册制度的效果,结果发现预注册可以提升研究质量,减少出版偏见,并提高研究的可信度^[123]。

5 同行评议的预见性

同行评议的预见性关注同行评议结果是否与之后的文献计量表现有关^[1]? 期刊或会议希望通过同行评议在提交的研究中遴选出最好的一部分,然而,这一目的的实现效果往往受到质疑^[124]。D. E. Chubin 等曾指出研究绩效和同行评议数据相结合可以揭示更多的规律^[125],因此,对文献计量和同行评议间的联系和差异的探讨成为分析同行评议预见性的重要切入点^[126]。

一方面,研究者讨论了文献计量与同行评议在遴选论文中的差异,如 L. Smolinsky 等研究了 MathSciNet 数据库中美国数学学会遴选出的优秀论文(featured review articles),发现 83.38% 优秀论文没有获得高被引,认为被引和专家评审有着弱相关关系^[127]。J. Du 等分析了 *F1000* 专家推荐与论文被引之间的关系,也总结出具有变革性意义的研究同行推荐多、被引低,循证研究被引高、同行推荐少的规律^[128]。J. Jirschitzka 等分析了跨学科会议的论文评分与被引间的关系,发现录用论文与会议的相关性(relevance to the conference)与被引频次呈正相关关系,而论文的新颖性则与被引频次负相关^[129]。

另一方面,研究者将同行评议因素与包括被引在内的文献计量指标相联系,谢维熙等利用 ICLR 公开评审数据分析了论文评审得分和被引频次间的一致性^[130]。如 J. Rigby 等发现经历审稿轮数更多的论文,被引相对更高,因此认为同行评议工作是有建设性的,即评审时被认为更需要修正的手稿,发表后被认为更有意义^[131]。而 S. Sikdar 等发现,对于被 *Journal of High Energy Physics* 录用的论文,经历更少轮评审的论文可能会得到更多的引用^[132]。审稿人网络还可以用于预测论文的影响力,如 S. Sikdar 等基于 *Journal of High Energy Physics* 的评审数据构建了审稿人网络,其

方法为,如果审稿人被同一个编辑指定,那么审稿人之间就形成一条边,而网络的结构特征(如度、集聚系数、中心性)则成为了预测论文被引量的有效指标^[132]。

调查被拒稿件的去向和影响力也是评估同行评议效果的常用方法,此类研究大多基于某本或几本期刊的评审数据,结果主要为被拒绝的稿件只有一部分最终发表了出来,并且发表在了影响因子相对更低的期刊上^[133]。然而,精英期刊则表现出了不同的特点,如 K. Siler 等统计了被 *BMJ*、*The Lancet* 等精英期刊拒稿的论文,发现有部分被拒稿的论文反而成为了高被引论文,这一点体现出了同行评议在论文质量把关中的瑕疵^[134]。基于预印本开展的研究也表现出了类似的特点,D. Y. Fu 等分析了 bioRxiv 上发布的预印本文章与其经过同行评议后的关系,发现有预印本的文章比没有预印本的文章平均高出 49% 的 Altmetrics 得分和 36% 的引用^[135]。F. Xu 等调查了发表在 bioRxiv 上的文章与其在期刊上发表后的一致性,发现与未在期刊上发表的预印本文章相比,在期刊上发表的文章在 Altmetrics 方面表现更好,而且预印本的 Altmetrics 和引用表现越好,其发表的期刊的影响因子也更高,这在一定程度上证明了同行评议的预见性^[136]。

除实验研究外,依靠观测数据和统计方法实施的“准实验”研究也在同行评议领域得以应用,此类研究观察“政策”实施后是否造成论文被引量的提升来评估“政策”的影响。刘瑞明等使用双重差分法证实了匿名审稿制度和名刊工程对国内经济学期刊和高校学报学术水平的促进作用^[137-138]。然而,OPR (Open peer review, 开放同行评议) 期刊样本量问题在一定程度上限制了以单个期刊样本所开展的研究结论的外推。以 OPR 和引用量的关系为例,J. Ni 等^[139]和 Q. Zong 等^[140]分别基于 *Nature Communications* 和 *PeerJ* 的同行评议数据集使用了匹配方法,但是前者发现 OPR 并未增加论文被引量,而后者却发现 OPR 提升了论文被引量。

6 结论与讨论

本文对科学计量视角下的同行评议研究进行了梳理,在概括同行评议量化特征的基础上,依托 L. Bornmann 所提出的研究框架,从同行评议公平性、可靠性、预见性 3 个方面对相关内容进行阐释,在此基础上得出结论及研究展望,以期学者开展后续研究提供参考,并为打开同行评议黑箱、提升同行评议质量提供一定借鉴。

从现有研究来看,数据的可获得性使得评审报告和审稿人成为研究者的重要关注点。在对审稿报告的研究中,研究者常从审稿意见长度和审稿速度切入展开分析,而随着自然语言处理技术的应用,研究者越发关注对审稿意见文本的挖掘。以往审稿人隐藏于幕后,随着数据的开放,揭示审稿人这一独特科研群体的身份特征引起了学界的兴趣,对审稿人的研究也延伸至对审稿人的评价与推荐。相比于初始的量化特征揭示,研究者进一步循着发现现象、揭示机理、提出和验证措施的路径进一步展开研究。在同行评议公平性方面,研究者关注同行评议中的性别、国家、机构和关系层面的不公平,并提出推行双盲审稿等制度来提升公平性。在可靠性方面,研究者从通过实证发现同行评议难以实现保证可靠到揭示可靠性背后的影响因素,并提出预注册等策略来提升同行评议的可靠性。在预见性方面,一些研究证实了同行评议的预见性,也有研究对此提出质疑,而研究者们也不再限于证明预见性,而是以预见性来检验期刊办刊水平的差异与变化,旨在揭示影响同行评议预见性的因素和检验提升预见性的举措。

总的来说,国际上较早对于同行评议开展计量研究,并且形成了以 F. Squazzoni 为核心的高产团队,该团队在研究方法多样化、数据的覆盖面、研究的影响力上都处于领先地位。相比之下,国内的相关研究仍在起步阶段,在该领域的产出和影响力仍相对较小,而且未形成高影响力作者群体。从研究上看,现有研究仍多以数据驱动的计量研究为主,较少关注研究的可解释性。因此,我们进一步从研究方法多样性、数据应用全面性和结论的可解释性 3 个方面展开讨论。

6.1 研究方法多样性

同行评议研究从最初的经验性阐释和描述统计发展到多种方法应用。以 ABM 为主的建模仿真方法曾在同行评议实证研究方面发挥过重要作用,值得一提的是,仿真的实施应当以事实为依据,并应探讨对同行评议的指导意义。随着数据不断开放,依托开放同行评议数据的数据驱动式的研究逐渐成为主流,网络分析研究更擅长揭示出同行评议中的社会关系及其学术影响,文本挖掘的使用则加深了对评议报告内在特征的理解。实验研究在同行评议研究中有着重要价值,特别是对期刊质量提升的探讨无法靠分析审稿报告来解决,因此期刊需要主动通过实验加以验证^[141]。

关于开放同行评议的实验研究,*BMJ* 在 1999 年设计随机对照实验后发现,公开审稿人身份并不会对评

审质量产生不利影响^[142]。2014 年,Elsevier 对 5 种期刊进行了从封闭同行评议转为开放同行评议的实验,发现开放评审报告同样不会影响同行评议过程^[19]。然而,考虑到实验的成本和提升质量的动机,同行评议领域的实验往往由出版商、会议和期刊组织并实施。也有学者在理论模型分析和博弈机制设计的基础上,在可控的范围内开展研究也可以在一定程度上揭示同行评议的内在机制,如 F. Squazzoni 等实施的“信任博弈”(Investment Game)实验^[143],通过招募学生开展实验,揭示了向审稿人提供物质奖励会降低评审过程的质量和效率^[144]。此外,尽管不是主流研究,数理模型层面的推演和论证^[145]也在分析同行评议中的互动机制上发挥着作用。

6.2 数据应用全面性

同行评议数据获取上的困难曾一度限制了研究的开展,而开放同行评议的发展为相关研究创造了一个有利的环境,使研究人员能够更快、更方便地获取同行评议数据^[141]。帮助审稿人跟踪同行评议活动的平台,例如 Publons、Peerage of Science,如今已成为研究者观察研究同行评议的重要窗口,其中 Publons 已成为同行评议研究者使用最多的数据源之一。就期刊、出版商和计算机会议而言,目前已有大量公开数据可供研究者使用^[146],比如 *Nature Communications*^[139]、*BMJ*^[36]、*PeerJ*^[2,140]、*BMC*^[51]、*F1000*^[128]。计算机领域的会议数据集包括 *PeerRead* 数据集^[43]和开放评审平台 *Openreview*^[130]。

然而高质量研究往往依赖大规模且高质量的数据,如以网络分析方法开展的研究不仅包含了期刊的录用论文数据,也包括拒稿论文的数据。这些数据往往由研究者与期刊或出版商通过签署协议获得^[100],而大部分 OPR 期刊仅开放了录用论文的信息。其次,相比于国际上的同行评议数据开放工作,国内期刊的举措相对保守。值得一提的是,《心理学报》等先锋期刊在其官网提供了匿名的论文评审报告^[107],已经朝着开放同行评议的方向做出了积极的尝试。除期刊和会议论文评审数据之外,美国数学学会出版的 *Mathematical Reviews* 等二次文献也为研究同行评议提供了素材^[127]。与之相似,国内人大书报资料中心所创办的复印报刊资料^[147-148]等二次文献也已被用于同行评议研究。此外,预印本亦成为观察和解读论文同行评议研究的数据源之一^[136]。

除直接的论文评审数据外,供研究人员分享经验的 *SciRev*^[47]以及带有后同行评议性质的学术论坛

PubPeer^[48]也已被研究者用于讨论同行评议的过程和影响。最后,研究者使用大部分期刊开放同行评议数据往往只限于录用论文的数据,若仅对这一类型数据展开分析,应当考虑“伯克森悖论”(Berkson's paradox),避免得出有偏差的结论。

6.3 结论的可解释性

基于同行评议数据开展的实证研究为同行评议的科学化提供了重要的证据,从科学计量学发现现象到为政策提供借鉴也成为了同行评议研究中的重要方向。尽管数据的开放大大促进了同行评议研究的推进,然而由于数据集(往往为某领域或某本期刊评审数据)的限制,大量数据驱动型的研究得出了不一致结论,这可能会限制结论的外推。如在审稿速度方面^[28-29]、不同性别下的审稿意见长度差异^[17,19]、审稿轮数与被引的关系^[51,132]等。

更重要的是,单纯的描述统计并未对在统计学意义上发现的现象进行逻辑学意义上的归因,没有基于科学社会学、科学哲学的理论基础提出具有说服力的解释或具有启发意义的见解。如果不能进一步挖掘同行评议机理,那么就难以为“打开同行评议黑箱”做出实质性的贡献。因此,针对同行评议开展的定量研究应当从描述统计过渡到挖掘同行评议背后的机理,从单纯的数据驱动式的研究过渡到开展更具可解释性的研究可能是未来同行评议研究中的重要方向之一。

参考文献:

- [1] BORNMAN L. Scientific peer review[J]. Annual review of information science and technology, 2011, 45(1): 197-245.
- [2] MATSUI A, CHEN E, WANG Y, et al. The impact of peer review on the contribution potential of scientific papers[J]. Peerj, 2021, 9: e11999.
- [3] FERREIRA C, BASTILLE-ROUSSEAU G, BENNETT A M, et al. The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline? [J]. Biological reviews, 2016, 91(3): 597-610.
- [4] GRIMALDO F, MARUŠIĆ A, SQUAZZONI F. Fragments of peer review: a quantitative analysis of the literature (1969-2015)[J]. Plos one, 2018, 13(2): e0193148.
- [5] DALTON M S. Refereeing of scholarly works for primary publishing [J]. Annual review of information science and technology, 1995, 30: 213-250.
- [6] 付伟棠. 我国学术期刊同行评议研究综述[J]. 中国科技期刊研究, 2019, 30(8): 819-826.
- [7] 常唯, 袁境泽. 国际学术出版中的同行评议进展与展望[J]. 中国科技期刊研究, 2020, 31(10): 1181-1192.
- [8] 索传军, 于淼. 国外期刊论文同行评议创新态势述评[J]. 图

- 书情报工作, 2021, 65(1): 128-139.
- [9] 秦成磊, 章成志. 大数据环境下同行评议面临的问题与对策[J]. 情报理论与实践, 2021, 44(4): 99-112.
- [10] BATAGELJ V, FERLIGOJ A, SQUAZZONI F. The emergence of a field: a network analysis of research on peer review[J]. *Scientometrics*, 2017, 113(1): 503-532.
- [11] RENNIE D. Let's make peer review scientific[J]. *Nature*, 2016, 535(7610): 31-33.
- [12] LABAND D N. Is there value-added from the review process in economics?: preliminary evidence from authors[J]. *Quarterly journal of economics*, 1990, 105(2): 341-352.
- [13] CASNICI N, GRIMALDO F, GILBERT N, et al. Attitudes of referees in a multidisciplinary journal: an empirical analysis[J]. *Journal of the Association for Information Science and Technology*, 2017, 68(7): 1763-1771.
- [14] BIANCHI F, GRIMALDO F, SQUAZZONI F. The F-3-index: valuing reviewers for scholarly journals[J]. *Journal of informetrics*, 2019, 13(1): 78-86.
- [15] LEE C J, SUGIMOTO C R, ZHANG G, et al. Bias in peer review[J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(1): 2-17.
- [16] ZHANG G, XU S, SUN Y, et al. Understanding the peer review endeavor in scientific publishing[J]. *Journal of informetrics*, 2022, 16(2): 101264.
- [17] BULJAN I, GARCIA-COSTA D, GRIMALDO F, et al. Large-scale language analysis of peer review reports[J]. *Elife*, 2020, 9: e53249.
- [18] ZHANG L, SHANG Y, HUANG Y, et al. Gender differences among active reviewers: an investigation based on Publons[J]. *Scientometrics*, 2022, 127(1): 145-179.
- [19] BRAVO G, GRIMALDO F, LÓPEZ-IÑESTA E, et al. The effect of publishing peer review reports on referee behavior in five scholarly journals[J]. *Nature communications*, 2019, 10(1): 322.
- [20] GARCIA-COSTA D, SQUAZZONI F, MEHMANI B, et al. Measuring the developmental function of peer review: a multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals[J]. *Ssrn*: 10.2139/ssrn.3912607, 2021.
- [21] BORNMANN L, WOLF M, DANIEL H D. Closed versus open reviewing of journal manuscripts: how far do comments differ in language use? [J]. *Scientometrics*, 2012, 91(3): 843-856.
- [22] CHETTY R, SAEZ E, SANDOR L. What policies increase prosocial behavior? an experiment with referees at the journal of public economics [J]. *Journal of economic perspectives*, 2014, 28(3): 169-188.
- [23] SALINAS S, MUNCH S B. Where should i send it? optimizing the submission decision process [J]. *Plos one*, 2015, 10(1): e0115451.
- [24] BILALLI B, MUNIR R F, ABELLÓ A. A framework for assessing the peer review duration of journals: case study in computer science [J]. *Scientometrics*, 2020, 126(1): 545-563.
- [25] XU S, AN M, AN X. Do scientific publications by editorial board members have shorter publication delays and then higher influence? [J]. *Scientometrics*, 2021, 126(8): 6697-6713.
- [26] SARIGOEL E, GARCIA D, SCHOLTES I, et al. Quantifying the effect of editor-author relations on manuscript handling times[J]. *Scientometrics*, 2017, 113(1): 609-631.
- [27] MROWINSKI M J, FRONCZAK A, FRONCZAK P, et al. Review time in peer review: quantitative analysis and modelling of editorial workflows[J]. *Scientometrics*, 2016, 107(1): 271-286.
- [28] DAVO M del C, VIVES C, ÁLVAREZ-DARDET C. Why are women underused in the JECH peer review process? [J]. *Journal of epidemiology & community health*, 2003, 57(12): 936-937.
- [29] SCHMALING K B, BLUME A W. Gender differences in providing peer review to two behavioural science journals, 2006-2015[J]. *Learned publishing*, 2017, 30(3): 221-225.
- [30] HUISMAN J, SMITS J. Duration and quality of the peer review process: the author's perspective[J]. *Scientometrics*, 2017, 113(1): 633-650.
- [31] PAUTASSO M, SCHÄFER H. Peer review delay and selectivity in ecology journals[J]. *Scientometrics*, 2010, 84(2): 307-315.
- [32] OKIKE K, KOCHER M S, NWACHUKWU B U, et al. The fate of manuscripts rejected by the journal of bone and joint surgery (American volume) [J]. *The journal of bone & joint surgery*, 2012, 94(17): e130.
- [33] DECULLIER E, LHÉRITIER V, CHAPUIS F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study[J]. *BMJ*, 2005, 331(7507): 19.
- [34] CHAREN D A, MAHER N A, ZUBIZARRETA N, et al. Evaluation of publication delays in the orthopedic surgery manuscript review process from 2010 to 2015[J]. *Scientometrics*, 2020, 124(2): 1127-1135.
- [35] HORBACH S P J M. Pandemic publishing: medical journals strongly speed up their publication process for COVID-19 [J]. *Quantitative science studies*, 2020, 1(3): 1056-1067.
- [36] DELGADO A F, GARRETSON G, FALK DELGADO A. The language of peer review reports on articles published in the BMJ, 2014-2017: an observational study[J]. *Scientometrics*, 2019, 120(3): 1225-1235.
- [37] ZHANG G, WANG L, XIE W, et al. "This article is interesting, however": exploring the language use in peer review comment of articles published in the BMJ[J]. *Aslib journal of information management*, 2022, 74(3): 399-416.
- [38] 张明阳, 王刚, 彭起, 等. 学术论文公开评审平台数据分析[J]. *计算机科学*, 2021, 48(6): 63-70.
- [39] HORBACH S P J M. No time for that now! qualitative changes in manuscript peer review during the Covid-19 pandemic [J]. *Research evaluation*, 2021, 30(3): 231-239.
- [40] HUA X, NIKOLOV M, BADUGU N, et al. Argument mining for understanding peer reviews[C]//2019 conference of the North American Chapter of the Association for Computational Linguistics:

- human language technologies. Minneapolis: Association for Computational Linguistics, 2019: 2131–2137.
- [41] QIN C, ZHANG C. Exploring the distribution regularities of referees' comments in IMRAD structure of academic articles[C]//18th International conference on scientometrics and informetrics conference. Leuven: International Society for Scientometrics and Informetrics, 2021: 1527–1528.
- [42] MENG M, WANG Y, ZHANG C. Building multi-level aspects of peer reviews for academic articles[C]//18th international conference on scientometrics and informetrics conference. Leuven: International Society for Scientometrics and Informetrics, 2021: 1519–1520.
- [43] KANG D, AMMAR W, DALVI B, et al. A dataset of peer reviews (PeerRead): collection, insights and NLP applications[C]//Proceedings of the 2018 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies. New Orleans: Association for Computational Linguistics, 2018: 1647–1661.
- [44] 林原, 王凯巧, 丁堃, 等. 学术论文的定性评价量化研究[J]. 情报理论与实践, 2021, 44(8): 28–34.
- [45] RASHIDI K, SOTUDEH H, MIRZABEIGI M, et al. Determining the informativeness of comments: a natural language study of F1000Research open peer review reports[J]. Online information review, 2020, 44(7): 1327–1345.
- [46] VINCENT-LAMARRE P, LARIVIÈRE V. Textual analysis of artificial intelligence manuscripts reveals features associated with peer review outcome[J]. Quantitative science studies, 2021, 2(2): 662–677.
- [47] JIANG S. Understanding authors' psychological reactions to peer reviews: a text mining approach[J]. Scientometrics, 2021, 126(7): 6085–6103.
- [48] BORDIGNON F. Self-correction of science: a comparative study of negative citations and post-publication peer review[J]. Scientometrics, 2020, 124(2): 1225–1239.
- [49] ORTEGA J L. The relationship and incidence of three editorial notices in pubpeer: errata, expressions of concern, and retractions[J]. Learned publishing, 2021, 34(2): 164–174.
- [50] ORTEGA J L. Classification and analysis of pubpeer comments: how a web journal club is used[J]. Journal of the Association for Information Science and Technology, 2022, 73(5): 655–670.
- [51] WOLFRAM D, WANG P, ABUZAHA F. An exploration of referees' comments published in open peer review journals: the characteristics of review language and the association between review scrutiny and citations[J]. Research evaluation, 2021, 30(3): 314–322.
- [52] LUO J, FELICIANI T, REINHART M, et al. Analysing sentiments in peer review reports: evidence from two science funding agencies[J]. Quantitative science studies, 2022, 2(4): 1271–1295.
- [53] 姜春林, 张立伟, 刘盛博. 图书情报学期刊“联锁编委”的社会网络分析[J]. 情报学报, 2014, 33(5): 481–490.
- [54] 张丽华, 曲建升. 期刊编委比非编委论文作者能更早探测出研究前沿吗[J]. 情报杂志, 2017, 36(8): 113–119.
- [55] YOUK S, PARK H S. Where and what do they publish? editors' and editorial board members' affiliated institutions and the citation counts of their endogenous publications in the field of communication[J]. Scientometrics, 2019, 120(3): 1237–1260.
- [56] 王兴. 国际学术期刊编委数量与科研产出评价指标的相关性研究——以经济学学科世界 984 所大学为例[J]. 重庆大学学报(社会科学版), 2017, 23(1): 61–70.
- [57] 卢小莉, 李晶, 吴登生. 基于期刊编委指数的科研机构学术影响力评价研究: 以地质学为例[J]. 情报学报, 2018, 37(1): 14–24.
- [58] XIE Y, WU Q, LI X. Editorial team scholarly index (ETSI): an alternative indicator for evaluating academic journal reputation[J]. Scientometrics, 2019, 120(3): 1333–1349.
- [59] 李江. 认可审稿人的学术贡献[J]. 图书情报知识, 2018(5): 2.
- [60] ORTEGA J L. Are peer-review activities related to reviewer bibliometric performance? a scientometric analysis of Publons[J]. Scientometrics, 2017, 112(2): 947–962.
- [61] 刘丽萍, 刘春丽. 基于 Publons 平台的审稿人贡献认可与评价研究[J]. 中国科技期刊研究, 2020, 31(1): 99–107.
- [62] 雷燕. 英文科技期刊学术影响力与审稿人学术及审稿表现关系研究与启示——Publons 医学领域期刊实证分析[J]. 中国科技期刊研究, 2021, 32(2): 206–213.
- [63] WEI C, BU Y, KANG L, et al. Directionality of paper reviewing and publishing of a scientist: a Granger causality inference[J]. Data science and informetrics, 2021, 1(1): 68–80.
- [64] SQUAZZONI F, BRAVO G, FARJAM M, et al. Peer review and gender bias: a study on 145 scholarly journals[J]. Science advances, 2021, 7(2): eabd0299.
- [65] FOX C W. Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution[J]. Scientometrics, 2017, 113(1): 465–477.
- [66] GARCIA J A, RODRIGUEZ-SÁNCHEZ R, FDEZ-VALDIVIA J. The interplay between the reviewer's incentives and the journal's quality standard[J]. Scientometrics, 2021, 126(4): 3041–3061.
- [67] COPIELLO S. On the money value of peer review[J]. Scientometrics, 2018, 115(1): 613–620.
- [68] BIANCHI F, SQUAZZONI F. Is three better than one? simulating the effect of reviewer selection and behavior on the quality and efficiency of peer review[C]//2015 winter simulation conference. New York: IEEE, 2015: 4081–4089.
- [69] KALMUKOV Y. An algorithm for automatic assignment of reviewers to papers[J]. Scientometrics, 2020, 124(3): 1811–1850.
- [70] CHUGHTAI G R, LEE J, SHAHZADI M, et al. An efficient ontology-based topic-specific article recommendation model for best-fit reviewers[J]. Scientometrics, 2020, 122(1): 249–265.

- [71] GARCIA J A, RODRIGUEZ-SÁNCHEZ R, FDEZ-VALDIVIA J. The author-reviewer game [J]. *Scientometrics*, 2020, 124 (3): 2409–2431.
- [72] FELICIANI T, LUO J, MA L, et al. A scoping review of simulation models of peer review [J]. *Scientometrics*, 2019, 121 (1): 555–594.
- [73] 梁玉成, 贾小双. 数据驱动下的自主行动者建模[J]. 贵州师范大学学报(社会科学版), 2016(6): 31–34.
- [74] SOBKOWICZ P. Innovation suppression and clique evolution in peer-review-based, competitive research funding systems: an agent-based model [J]. *Journal of artificial societies and social simulation*, 2015, 18(2): 13.
- [75] RIGHI S, TAKÁCS K. The miracle of peer review and development in science: an agent-based model [J]. *Scientometrics*, 2017, 113(1): 587–607.
- [76] BIANCHI F, GRIMALDO F, BRAVO G, et al. The peer review game: an agent-based model of scientists facing resource constraints and institutional pressures [J]. *Scientometrics*, 2018, 116 (3): 1401–1420.
- [77] SQUAZZONI F, GANDELLI C. Saint Matthew strikes again: an agent-based model of peer review and the scientific community structure [J]. *Journal of informetrics*, 2012, 6(2): 265–275.
- [78] BAUTISTA CABOTA J, GRIMALDO F, SQUAZZONI F. When competition is pushed too hard. an agent-based model of strategic behaviour of referees in peer review [C]// Proceedings 27th european conference on modelling and simulation. Nottingham: European council modeling & simulation, 2013.
- [79] MARSH H W, BORNMAN L, MUTZ R, et al. Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches [J]. *Review of educational research*, 2009, 79(3): 1290–1326.
- [80] PINHO-GOMES A C, VASSALLO A, THOMPSON K, et al. Representation of women among editors in chief of leading medical journals [J]. *Jama network open*, 2021, 4(9): e2123026.
- [81] FOX C W, DUFFY M A, FAIRBAIRN D J, et al. Gender diversity of editorial boards and gender differences in the peer review process at six journals of ecology and evolution [J]. *Ecology and evolution*, 2019, 9(24): 13636–13649.
- [82] HELMER M, SCHOTTDORF M, NEEF A, et al. Gender bias in scholarly peer review [J]. *Elife*, 2017, 6: e21718.
- [83] SQUAZZONI F, BRAVO G, GRIMALDO F, et al. Gender gap in journal submissions and peer review during the first wave of the covid-19 pandemic. a study on 2329 Elsevier journals [J]. *Plos one*, 2021, 16(10): e0257919.
- [84] MURRAY D, SILER K, LARIVIÈRE V, et al. Gender and international diversity improves equity in peer review [J]. *Biorxiv*: 10.1101/400515, 2018.
- [85] DEMAREST B, FREEMAN G, SUGIMOTO C R. The reviewer in the mirror: examining gendered and ethnicized notions of reciprocity in peer review [J]. *Scientometrics*, 2014, 101(1): 717–735.
- [86] FOX C W, PAINE C E T. Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution [J]. *Ecology and evolution*, 2019, 9(6): 3599–3619.
- [87] CARD D, DELLAVIGNA S, FUNK P, et al. Are referees and editors in economics gender neutral? [J]. *The quarterly journal of economics*, 2020, 135(1): 269–327.
- [88] FOX C W, BURNS C S, MEYER J A. Editor and reviewer gender influence the peer review process but not peer review outcomes at an ecology journal [J]. *Functional ecology*, 2016, 30 (1): 140–153.
- [89] BURNS C S, FOX C W. Language and socioeconomics predict geographic variation in peer review outcomes at an ecology journal [J]. *Scientometrics*, 2017, 113(2): 1113–1127.
- [90] ZHU H. Home country bias in academic publishing: a case study of the new England journal of medicine [J]. *Learned publishing*, 2021, 34(4): 578–584.
- [91] ZHANG X. Effect of reviewer's origin on peer review: China vs. non-China [J]. *Learned publishing*, 2012, 25(4): 265–270.
- [92] CAMPOS-ARCEIZ A, PRIMACK R B, KOH L P. Reviewer recommendations and editors' decisions for a conservation journal: is it just a crapshoot? and do Chinese authors get a fair shot? [J]. *Biological conservation*, 2015, 186: 22–27.
- [93] WEI Y, LEI L. Institution bias in the new England journal of medicine? a bibliometric analysis of publications (1997–2016) [J]. *Scientometrics*, 2018, 117(3): 1771–1775.
- [94] 郭峰, 李欣. 编辑部偏爱、关系稿与引用率贴水——来自中国经济学权威期刊的证据 [J]. *经济学(季刊)*, 2017, 16(4): 1237–1260.
- [95] BLANK R. The effects of double-blind versus single-blind reviewing- experimental-evidence from the American-economic-review [J]. *American economic review*, 1991, 81(5): 1041–1067.
- [96] TOMKINS A, ZHANG M, HEAVLIN W D. Reviewer bias in single-versus double-blind peer review [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(48): 12708–12713.
- [97] SEEGER M, BACCHELLI A. Does single blind peer review hinder newcomers? [J]. *Scientometrics*, 2017, 113(1): 567–585.
- [98] FOX C W, BURNS C S, MUNCY A D, et al. Author-suggested reviewers: gender differences and influences on the peer review process at an ecology journal [J]. *Functional ecology*, 2017, 31(1): 270–280.
- [99] SHOPOVSKI J, BOLEK C, BOLEK M. Characteristics of peer review reports: editor-suggested versus author-suggested reviewers [J]. *Science and engineering ethics*, 2020, 26(2): 709–726.
- [100] TEPLITSKIY M, ACUNA D, ELAMRANI-RAOULT A, et al. The sociology of scientific validity: how professional networks shape judgement in peer review [J]. *Research policy*, 2018, 47(9): 1825–1841.
- [101] BRAVO G, FARJAM M, GRIMALDO MORENO F, et al. Hidden connections: network effects on editorial decisions in four

- computer science journals[J]. Journal of informetrics, 2018, 12 (1): 101–112.
- [102] MARSH H W, JAYASINGHE U W, BOND N W. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability[J]. American psychologist, 2008, 63 (3): 160–168.
- [103] CICCHETTI D V. The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation[J]. Behavioral and brain sciences, 1991, 14(1): 119–135.
- [104] BORNHANN L, MUTZ R, DANIEL H D. A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants [J]. Plos one, 2010, 5 (12): e14331.
- [105] 王颖, 孔爱英, 朱蓓, 等. 科技期刊审稿标准一致性的影响因素及对策[J]. 中国科技期刊研究, 2014, 25 (12): 1470–1472.
- [106] ROTHWELL P M, MARTYN C N. Reproducibility of peer review in clinical neuroscience-is agreement between reviewers any greater than would be expected by chance alone? [J]. Brain, 2000, 123(9): 1964–1969.
- [107] 张光耀, 谢维熙, 夏鑫璐, 等. 非共识研究的影响力更高还是更低呢? ——基于中文开放同行评议审稿意见的分析[EB/OL]. [2022-06-21]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20220329.1252.002.html>
- [108] 刘欣, 李江, 吴金闪. 同行评议一致性研究[J]. 信息资源管理学报, 2022, 11(6): 10–16.
- [109] 岳名亮, 李富山, 汤宏波, 等. 期刊审稿专家一致性评价方法及其有效性验证[J]. 数据分析与知识发现, 2021, 5(4): 115–122.
- [110] YUE M, TANG H, LIU F, et al. Consistency index: measuring the performances of scholar journal reviewers[J]. Scientometrics, 2021, 126(8): 7183–7195.
- [111] KRAVITZ R L, FRANKS P, FELDMAN M D, et al. Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? [J]. Plos one, 2010, 5 (4): e10072.
- [112] FYFE A, SQUAZZONI F, TORN Y D, et al. Managing the growth of peer review at the royal society journals, 1865–1965 [J]. Science technology & human values, 2020, 45(3): 405–429.
- [113] RESNIK D B, ELMORE S A. Ensuring the quality, fairness, and integrity of journal peer review: a possible role of editors [J]. Science and engineering ethics, 2016, 22(1): 169–188.
- [114] PETERSEN A M. Megajournal mismanagement: manuscript decision bias and anomalous editor activity at plos one [J]. Journal of informetrics, 2019, 13(4): 100974.
- [115] PETERS D P, CECI S J. Peer-review practices of psychological journals: the fate of published articles, submitted again[J]. Behavioral and brain sciences, 1982, 5(2): 187–195.
- [116] FRANCOIS O. Arbitrariness of peer review: a bayesian analysis of the NIPS experiment[J]. Arxiv:1507.06411. 2015.
- [117] BREZIS E S, BIRUKOU A. Arbitrariness in the peer review process[J]. Scientometrics, 2020, 123(1): 393–411.
- [118] CORTES C, LAWRENCE N D. Inconsistency in conference peer review: revisiting the 2014 NeurIPS experiment [J]. Arxiv: 2109.09774. 2021.
- [119] SEEBER M. How do journals of different rank instruct peer reviewers? reviewer guidelines in the field of management[J]. Scientometrics, 2020, 122(3): 1387–1405.
- [120] SAKAI D. Who is peer reviewed? comparing publication patterns of peer-reviewed and non-peer-reviewed papers in Japanese political science[J]. Scientometrics, 2019, 121(1): 65–80.
- [121] HORBACH S P J M, HALFFMAN W. The ability of different peer review procedures to flag problematic publications[J]. Scientometrics, 2019, 118(1): 339–373.
- [122] LEEK J T, TAUB M A, PINEDA F J. Cooperation between referees and authors increases peer review accuracy [J]. Plos one, 2011, 6(11): e26895.
- [123] SODERBERG C, ERRINGTON T, SCHIAVONE S, et al. Initial evidence of research quality of registered reports compared with the standard publishing model [J]. Nature human behaviour, 2021, 5(8): 1–8.
- [124] SMITH R. Peer review: a flawed process at the heart of science and journals [J]. Journal of the Royal Society of Medicine, 2006, 99(4): 178–182.
- [125] CHUBIN D E, HACKETT E J. Peerless science: peer review and U.S. science policy [M]. New York: State University of New York Press, 1990.
- [126] ABRAMO G, D'ANGELO C A, REALE E. Peer review versus bibliometrics: which method better predicts the scholarly impact of publications? [J]. Scientometrics, 2019, 121 (1): 537–554.
- [127] SMOLINSKY L, SAGE D S, LERCHER A J, et al. Citations versus expert opinions: citation analysis of featured reviews of the American Mathematical Society [J]. Scientometrics, 2021, 126 (5): 3853–3870.
- [128] DU J, TANG X, WU Y. The effects of research level and article type on the differences between citation metrics and F1000 recommendations[J]. Journal of the Association for Information Science and Technology, 2016, 67(12): 3008–3021.
- [129] JIRSCHITZKA J, OEERST A, GÖLLNER R, et al. Inter-rater reliability and validity of peer reviews in an interdisciplinary field [J]. Scientometrics, 2017, 113(2): 1059–1092.
- [130] 谢维熙, 张光耀, 王贤文. 开放同行评议视角下学术论文同行评议得分与被引频次的关系[J]. 中国科技期刊研究, 2022, 33(1): 113–121.
- [131] RIGBY J, COX D, JULIAN K. Journal peer review: a bar or bridge? an analysis of a paper's revision history and turnaround time, and the effect on citation[J]. Scientometrics, 2018, 114 (3): 1087–1105.

[132] SIKDAR S, MARSILI M, GANGULY N, et al. Influence of reviewer interaction network on long-term citations: a case study of the scientific peer-review system of the journal of high energy physics [C]//2017 ACM/IEEE joint conference on digital libraries. Toronto: IEEE, 2017: 1–10.

[133] CASNICI N, GRIMALDO F, GILBERT N, et al. Assessing peer review by gauging the fate of rejected manuscripts: the case of the journal of artificial societies and social simulation[J]. Scientometrics, 2017, 113(1): 533–546.

[134] SILER K, LEE K, BERO L. Measuring the effectiveness of scientific gatekeeping[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112(2): 360–365.

[135] FU D Y, HUGHEY J J. Releasing a preprint is associated with more attention and citations for the peer-reviewed article[J]. Elife, 2019, 8: e52646.

[136] XU F, OU G, MA T, et al. The consistency of impact of preprints and their journal publications[J]. Journal of informetrics, 2021, 15(2): 101153.

[137] 刘瑞明, 赵仁杰. 匿名审稿制度推动了中国的经济学进步吗?——基于双重差分方法的研究[J]. 经济学(季刊), 2017, 16(1): 173–204.

[138] 刘瑞明, 赵仁杰. 政府支持、制度变革与学术期刊进步——来自中国“名刊工程”的经验证据[J]. 经济学(季刊), 2020, 19(2): 473–498.

[139] NI J, ZHAO Z, SHAO Y, et al. The influence of opening up peer review on the citations of journal articles[J]. Scientometrics, 2021, 126(12): 9393–9404.

[140] ZONG Q, XIE Y, LIANG J. Does open peer review improve citation count? evidence from a propensity score matching analysis of PeerJ[J]. Scientometrics, 2020, 125(1): 607–623.

[141] SQUAZZONI F, AHRWEILER P, BARROS T, et al. Unlock ways to share data on peer review[J]. Nature, 2020, 578(7796): 512–514.

[142] VAN ROOYEN S, DELAMOTHE T, EVANS S J W. Effect on peer review of telling reviewers that their signed reviews might be posted on the Web: randomised controlled trial[J]. BMJ, 2010, 341: e5729.

[143] BERG J, DICKHAUT J, MCCABE K. Trust, reciprocity, and social-history[J]. Games and economic behavior, 1995, 10(1): 122–142.

[144] SQUAZZONI F, BRAVO G, TAKACS K. Does incentive provision increase the quality of peer review? an experimental study[J]. Research policy, 2013, 42(1): 287–294.

[145] RODRIGUEZ-SÁNCHEZ R, GARCÍA J A, FDEZ-VALDIVIA J. Editorial decisions with informed and uninformed reviewers[J]. Scientometrics, 2018, 117(1): 25–43.

[146] 彭琳, 杜杏叶. 科技期刊实施开放式同行评议策略研究[J]. 中国科技期刊研究, 2018, 29(11): 1114–1121.

[147] 姜春林, 张光耀, 郭琪琴. 复印报刊资料科技哲学卷文献计量分析[J]. 自然辩证法研究, 2019, 35(12): 69–75.

[148] 姜春林, 贾龙川, 张光耀. 人大复印报刊资料《创新政策与管理》文献计量研究[J]. 山东科技大学学报(社会科学版), 2020, 22(1): 31–39.

作者贡献说明:

张光耀:提出研究思路,论文撰写和论文修改;
谢维熙:文献收集和论文修改;
姜春林:论文修改;
王贤文:提出研究思路,论文修改与审定。

A Review of Peer Review of Papers from the Scientometrics Perspective

Zhang Guangyao^{1,2} Xie Weixi¹ Jiang Chunlin¹ Wang Xianwen¹

¹ Institute of Science of Science and S&T Management and WISE Lab,
Dalian University of Technology, Dalian 116024

² Maastricht Economic and Social Research Institute on Innovation and Technology,
United Nations University, Maastricht 6211AX

Abstract: [Purpose/Significance] By combing related researches of peer review from the scientometrics perspective, this paper provides references for further in-depth research on peer review. [Method/Process] This paper reviewed the current foreign and domestic status of research from three aspects: fairness, reliability, and validity of peer review on the basis of generalizing the quantitative characteristics of peer review. [Result/Conclusion] The paper summarizes the trends in the diversity of research methods and the comprehensiveness of data application in peer review research from the scientometrics perspective, and points out that the interpretability of the current findings still needs to be improved, and the empirical research on peer review in China is still relatively insufficient.

Keywords: peer review open peer review science of science scientometrics