

- 张光耀, 谢维熙, 夏鑫璐, 耿 屿, 姜春林, 王贤文 (大连理工大学科学学与科技管理研究所暨 WISE 实验室, 辽宁 大连 116024)

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

摘 要: [目的/意义] 同行评议中的非共识备受学术界关注。在开放同行评议的背景下, 探索同行评议中的非共识度与论文影响力间的关系。[方法/过程] 以国内开放同行评议期刊《心理学报》的开放审稿意见为研究对象, 构建了2014—2020年发表的888篇论文的22604条有效评审意见的数据集, 设计审稿意见非共识度指标, 采用回归分析检验审稿意见非共识度和论文影响力间的关系。[结果/结论] 审稿意见非共识度与论文影响力存在统计上的相关关系, 在加入控制变量后, 进一步验证审稿意见非共识度与论文影响力之间存在显著的负向关系, 为理解同行评议提供了新的视角。

关键词: 同行评议; 非共识; 开放同行评议; 科学学 《心理学报》

DOI: 10.16353/j.cnki.1000-7490.2022.09.014

引用格式: 张光耀, 谢维熙, 夏鑫璐, 耿屿, 姜春林, 王贤文. 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 [J]. 情报理论与实践, 2022, 45 (9): 102-109.

Do Non-Consensus Studies Have Higher or Lower Impact? Analysis Based on Chinese Open Review Comments

Abstract [Purpose/significance] The non-consensus in peer review has attracted the attention of academia. This research intends to explore the relationship between non-consensus in peer review and papers' impact in the context of open peer review. [Method/process] Taking the open review comments of "Acta Psychologica Sinica" as the research object, based on the 22604 peer review records of 888 papers published between 2014 and 2020, the review non-consensus index was designed, and regression analysis was used to test the relationship between the review non-consensus index and the papers' impact separately. There is a statistical correlation between the review diversity and the papers' impact. [Result/conclusion] After adding the control variables, it was found that there was a significantly negative relationship between the review non-consensus index and papers' impact. This research is the first domestic preliminary exploration of peer review based on Chinese open review comments, and provides a new perspective for understanding peer review.

Keywords: peer review; non-consensus; open peer review; science of science; Acta Psychologica Sinica

同行评议是科学发展的基石^[1]。长期以来, 同行评议在质量控制、学术交流和提升论文质量等方面的重要作用得到了普遍认可, 但也面临诸多争议和批评^[2-3]。在同行评议所面对的众多质疑中, “非共识”备受关注, 评审的非共识现象广泛存在于项目评审和论文评审中。虽然国内曾有学者对项目评审中的非共识现象加以探讨^[4-5], 但这一重要现象在论文同行评议研究中鲜有提及。在这一现实背景之下, 随着开放同行评议运动的不断推进, 学术界开启了对同行评议中论文评审的非共识现象的探究。本文将在梳理同行评议非共识现象的基础上, 基于国内最早实行

开放同行评议期刊《心理学报》的开放审稿意见, 设计审稿意见非共识度指标, 检验非共识度和论文影响力之间的关系, 从而为揭示非共识背后的科学学意涵进行有价值的探讨。

1 文献综述

1.1 研究背景: 同行评议中的非共识

杨列勋等^[5]在对项目同行评议的研究中对非共识的定义是: 对于一个项目的评审结果, 如果有部分专家持肯定评价, 另一部分专家持否定评价, 且各有充分依据, 这类项目则被视作非共识项目。在评审过程中, 评审专家对同一评审项目的科学意义、创新性、可行性以及是否给予资助或录用时常持有相左的观点^[6]。与此同时, 非共识在

^{*} 本文为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”的研究成果, 项目编号: 71974029。

论文评审中广泛存在。已有研究指出,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^[7],来自国际放射学^[8]、神经科学^[9]和康复学^[10]的证据也表明论文评审意见不一致现象时常发生。

学术界对非共识的态度比较复杂,非共识常被认为是同行评议难以有效运行的表现^[11]。非共识的产生原因是多样的,从认识论来看,尽管评审专家力求采用同一套标准对科学发现进行评价,然而同行评议作为决策过程,是多元主体共同作用的结果,因此难免对新观点或新理论产生不一致的看法,即出现“非共识”^[4-6]。从以审稿人为主体的研究来看,不同类型的审稿人在评审活动中存在差异,如初级学者比资深学者的审稿速度更快,审稿意见更细致,而资深学者虽然比初级学者审稿量小,但拒审次数更多^[12-13]。与此同时,审稿人的非理性或者偏见,审稿人专业经验和专业对口程度都可能对审稿结果产生影响,且由于首因效应、晕轮效应等存在,审稿过程难以避免认知偏差。除此之外,还有另一条解释路径是,非共识源自审稿人对创新的接受程度不足,甚至认为“非共识”可能在一定程度上代表着创新性,从而提出同行评议有可能否定重大创新观点^[4,14]。如很多学者在提及非共识现象时常引用阿克洛夫和格兰诺维特等学者的案例^[15-16],大量学者也由此认为非共识可能来源于科学时代的局限性,但是这种案例可能只是科学史中的少数,只是由于影响巨大而获得学术界大量关注^[17]。

然而,也有研究者指出,追求一致性或许是不利于同行评议的。具体而言,高一致性实际上表明评审进程运作不良,没有体现审稿人意见的多样性,出现了审稿冗余^[18]。换言之,审稿人的一致性低并不代表评审的有效性低;反之,这可能表明审稿人是高度胜任的,因为它体现出了对“好的”和“有价值的”研究的各种相对中立的观点^[19]。与此同时,争论和分歧对科学是有益的,审稿人的意见不一致,并不代表审稿人无法对“好科学”达成一致^[20]。

实际上,依照科学社会学家默顿的观点,“科研成果由其内在价值衡量”^[21]。科研成果的影响力可能在其评审过程中得以体现。如有实证研究表明,经过多学科专家评审并推荐发表的文章比同一学科的专家评审的文章更有可能得到引用^[13]。从文章的价值水平来看,科学中更为常见的不是具有重大创新的研究,而是创新程度有限的研究^[5]。因此,非共识不一定是同行评议有效性低的体现,也可能意味着研究的学术价值有限。换言之,该类论文可能有着相对较低的影响力。

目前,学界对论文同行评议中非共识的研究还停留在较为表浅的层面,主要集中于对非共识程度的争论的层

面^[22],而往往忽视非共识背后的科学学意义。本文研究认为,尽管多数研究把低共识当作同行评议有效性不足的证据,但审稿人的非共识度也为理解同行评议提供了新视角。因此,本文将集中关注非共识的科学学意义,特别是在研究绩效有关的数据和同行评议数据相结合的情境下^[14],其中引用是最常被使用的衡量论文影响力的指标,为科学中的信任和注意力流动提供了一个客观和量化的衡量标准^[23]。因此,本文探讨的问题是:论文的非共识和论文影响力是否存在负向关系?

1.2 数据背景: 开放同行评议

尽管非共识广受作者和期刊的关注,也有着重要的研究价值,但由于传统同行评议的封闭性,对这一问题的研究局限于编辑部的自身经验^[24]。随着开放运动的不断深入^[25]，“British Medical Journal”“Nature Communications”等期刊纷纷开放了审稿意见。研究者可以打开同行评议黑箱,从微观层面基于期刊审稿文本开展研究工作^[26]。秦成磊等^[27]分析审稿意见在学术论文不同结构功能中的分布。林原等^[28]基于 ICLR 开放评审数据,提出了融合 Twitter 文本情感特征和同行评议情感特征来推进进行定性评价量化。徐申萌等^[29]基于 Publons 提供的数据研究了审稿人和审稿记录在世界的分布情况,并提出将审稿意见长度作为反映审稿人在审稿过程中花费时间和精力潜在指标以理解审稿人的行为。Black 等^[30]经过实证分析指出审稿人专业、高职称、学术地位与审稿质量关系不大,而审稿人经历的流行病学和统计学训练、年龄与审稿质量密切相关。Ni 等^[31]曾以“Nature Communications”为例,探讨了开放同行评议对论文被引的影响,结果发现开放评审报告没有造成论文被引量上升。国内期刊的开放同行评议进程仍在起步阶段,《心理学报》是尝试开放同行评议意见较早的期刊。《心理学报》的努力无疑是朝着开放同行评议迈出了一大步,也为本文探讨审稿人非共识现象提供了机会。本文将基于《心理学报》所提供的审稿意见文本,依据审稿意见数量和长度分别设计了非共识度指标,并通过探索非共识度指标和论文被引间的关系,为理解同行评议行为和进一步为审稿体系的规范化建设提供一定的量化依据。

2 数据和研究设计

2.1 数据

《心理学报》由中国心理学会和中国科学院心理研究所主办,是国内学术影响力较大的心理学期刊,在中国心理学界享有较高的声誉^[32]。为了推动期刊评议体系的不断完善,2014 年 3 月 11 日,《心理学报》编辑部决定公开评议专家的审稿意见。诚然,《心理学报》只是进行了

部分开放同行评议,但其已经向开放同行评议的方向迈出了一大步,并在促进学术交流公开透明和营造公平公开的学术环境方面取得了明显成效^[32]。因此,作为研究对象具有一定的代表性。

本文从《心理学报》期刊网站获取了2014—2020年发表的888篇论文的审稿意见信息,包括审稿轮数、审稿人标识,以及论文标题、作者在内的题录信息,并利用论文标题从中国知网匹配论文的引文量,作为分析样本(数据采集日期2020年10月),建立了22604条有效评审意见的数据集。

2.2 研究设计

考虑到不同轮数之间审稿意见的差异性,而且第一轮审稿往往是最重要的环节^[33-34]。因此,本文的研究只考虑第一轮13824条审稿意见。由于审稿意见的非共识的测度是较为主观的,涉及审稿人的态度、审稿意见的重要程度、评价用语的晦涩程度等要素,此外Luo等的研究也说明了情绪分析在同行评议报告中的不可靠性^[35]。因此,本文借鉴了Squazzoni等将审稿意见数量和长度作为测量审稿人的评审信息量的潜在指标的思路^[13,34],审稿意见长度的重要性在Ni等^[31]和张琳等^[36]的研究中也得到认可。尽管审稿意见的数量和长度差异仅仅反映了审稿人非共识的一个侧面,但仍然不失为研究同行评议非共识的有价值的切入点。

更重要的是,《心理学报》所提供的审稿意见是逐条列出的,因此不需要进行人工或机器的判断即可直接用于研究。与此同时,在设计非共识度的反映指标方面,郑兴东等^[37]在测量评审专家的非共识中曾使用专家评分的多项平均绝对离差和平均相对离差,杨列勋等^[5]曾采用评分方差来评估非共识研究项目,刘欣也曾以ILCR会议的评分来测度论文评审人评分的一致性^[38]。本文在前人研究的基础上发展了非共识度指标,引入了非共识度(数量)(ncon_num)和非共识度(长度)(ncon_len)两种指标。相比于非共识度(数量),非共识度(长度)在计算过程中损失了更多信息,因此本文则将后者用于对前者的稳健性检验。如果非共识度指标越接近0,则说明评审专家的非共识越小;如果指标越大,则说明专家们的非共识越大。具体计算见公式(1)和公式(2)。

非共识度(数量):

$$ncon_num = \sum (X_i - X_m)^2 / N \quad (1)$$

式中,ncon_num从审稿人的审稿意见的数量差异出发测度非共识,其中 X_i 代表每个审稿人 i 的审稿意见数量; X_m 代表一篇文章中每个审稿人审稿意见数量的平均值; N 代表每篇文章的审稿人数量。

非共识度(长度):

$$ncon_len = \sum (X_j - X_n)^2 / (10000 \times N) \quad (2)$$

式中,ncon_len从审稿人的审稿意见的长度差异出发测度非共识, X_j 代表每个审稿人 j 的所有审稿意见总长度; X_n 代表一篇文章中每个审稿人审稿意见的平均长度; N 代表每篇文章的审稿人数量。

3 结果分析

3.1 审稿意见量化特征描述

本文首先揭示了审稿意见数量和长度的量化特征,图1(a)和图1(b)分别呈现了审稿人的审稿意见数量和审稿意见长度分布,可以看出二者有着明显的右偏分布特征,审稿人撰写意见的中位数是7条,长度的中位数是710字。图1(c)和图1(d)呈现了审稿人平均意见数量和平均长度在不同轮数下的差异。第1轮的平均意见数量和长度最高,而且随着审稿轮数的深入,审稿人的平均意见数量和平均意见长度都呈现下降趋势。

本文对非共识度指标的数据特征进行进一步分析。如上所述,当涉及多轮审稿意见时,首轮审稿意见通常是评语最长且意见数量最多的。不同论文的审稿轮数差别很大,如果无差别考虑全部论文所有轮数的审稿意见,可能得出不准确的研究结果。因此,在计算非共识度时,仅采用了第1轮的审稿意见数据,并删除了只有一个审稿人的条目。图2(a)和图2(b)呈现了非共识度(数量)与非共识度(长度)。总体而言,每篇文章的审稿人从2~5不等,以2人和3人为主,分别占比为71.95%和26.34%。审稿意见非共识度(数量)的中位数为2.25,非共识度(长度)的中位数为3.82。随着审稿人数量的增加,两种非共识度指标都呈现上升趋势。最后,采用分仓散点图呈现非共识度和论文被引频次之间的关系,二者呈现出负相关关系。

3.2 回归分析

描述性统计数据表明,非共识度和论文影响力之间存在负相关关系。为了探索这种关系在控制其他因素时是否仍然成立,原假设为:

H0: 非共识度越高的论文影响力低。

备择假设为:

H1: 论文的非共识度和论文影响力之间没有统计上的显著相关性。

回归模型为

$$Y_i = x_i \beta + \varepsilon_i \quad (i = 1, \dots, n)$$

式中, Y_i 代表论文的被引量; x_i 为自变量和控制变量; ε_i

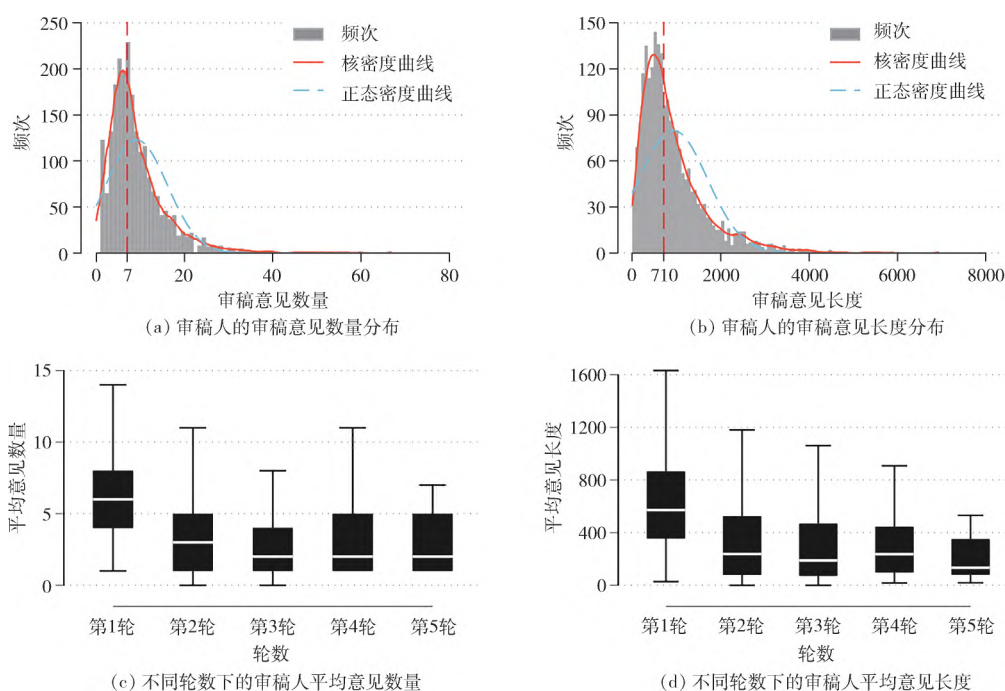


图1 审稿人的审稿意见数量和长度分布

Fig. 1 Distribution of the number and length of reviewers' comments

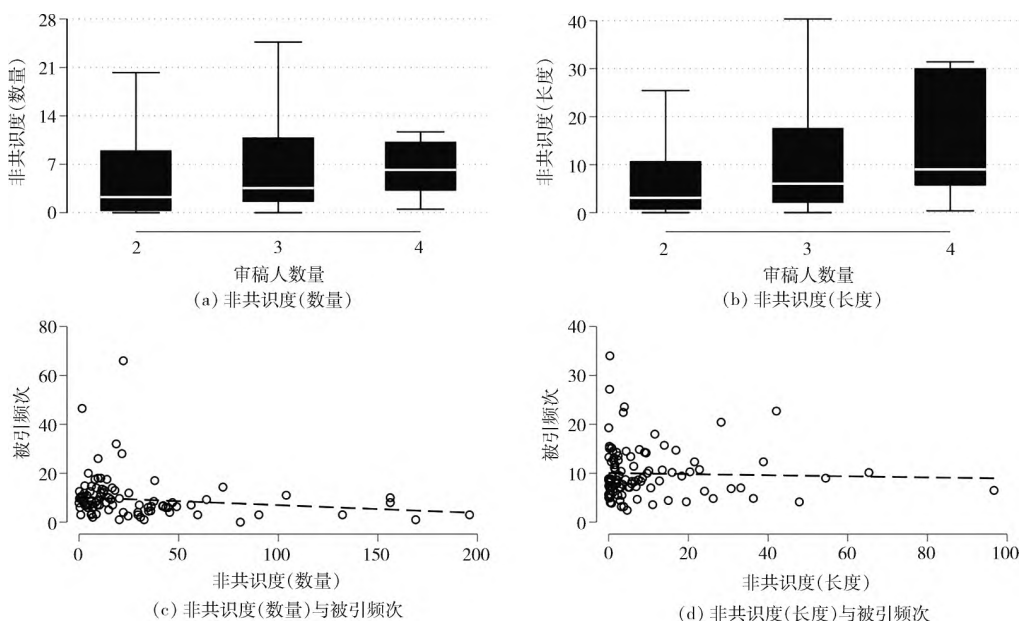


图2 非共识度分布特征及与被引频次的散点图

Fig. 2 Non-consensus distribution and scatterplot of non-consensus and citations

为误差项。表1给出了模型所使用的变量。

因变量：被引量（论文发表后当年及两年后的被引量）。

自变量：从两个方面来衡量论文的非共识特征，分别为非共识度（数量）和非共识度（长度）。

控制变量：在以往的研究基础上，将一些可能与论文

影响力有关的作者相关因素和论文相关因素作为控制变量，最终确定审稿人数量（re_num）、参考文献数量（ref）、页数（page）、作者数量（au_num）、基金资助（fund）、机构数量（org_num）、机构级别（org_rank）为控制变量。其中，审稿人数量获取自审稿意见，机构数量的计算为论文机构信息中不同单位的数量，并按最高一级单位统计。机构级别的计算为论文机构信息中第一作者所属机构的层次，编码为一流机构（一流机构包括“985”高校和中国科学院）或一般机构（具体的变量设定见表1）。此外，本文还控制了年份变量。从表2可知，描述性和相关性统计表明变量间有着较低的相关性，而且将变量加入回归后，VIF的最大值分别为1.26，说明不存在严重的共线性。

由于被引频次是离散变量，针对这一类变量，负二项模型是更常用的选择。本文

将自变量进行标准化处理以消除量纲后，采用负二项模型进行分析。表3汇报了纳入自变量和控制变量后的基础回归结果，其中， m_1 中只考虑了非共识度， m_2 中加入了审稿人数量和年份虚拟变量， m_3 中加入了其他控制变量。总体而言，非共识度和论文影响力呈负相关关系，而且在加入其他控制变量后，仍在0.05水平上显著。因此，假

设初步得到验证。

为了得到更加稳健的结果,表4进一步展开了稳健性检验。首先,考虑了结果对模型的依赖, $m4$ 中给出了Poisson回归结果,如果根据“更高的loglikelihood值、更

表1 变量及描述
Tab.1 Variable description

变量	符号	类型	描述
因变量	cit	count	发表当年及后两年被引量
	cit_cum	count	发表后累积被引量
自变量	ncon_num	count	非共识度(数量)
	ncon_len	count	非共识度(长度)
控制变量	re_num	dummy	审稿人数量
	ref	count	参考文献数量
	page	count	页数
	au_num	count	作者数量
	fund	dummy	是否为国家级基金资助
	org_num	count	机构数量
	org_rank	dummy	机构级别
	year	dummy	年份

表2 变量的描述统计结果和相关系数矩阵($N=669$)

Tab.2 Summary statistics and correlation matrix of variables ($N=669$)

变量	M	SD	1	2	3	4	5	6	7	8	9	10	11
cit	9.91	12.24	1										
cit_cum	20.67	42.80	0.77	1									
ncon_num	8.76	18.81	-0.05	-0.04	1								
ncon_len	11.68	27.93	-0.03	-0.03	0.51	1							
re_num	2.29	0.48	0.05	0.05	0.05	0.02	1						
ref	46.40	18.56	0.30	0.16	0.03	0.07	-0.00	1					
page	11.55	2.36	0.17	0.16	0.07	0.11	-0.05	0.50	1				
au_num	3.79	1.55	-0.11	-0.12	0.03	0.01	0.03	0.01	0.01	1			
fund	0.70	0.46	0.02	0.06	0.03	0.02	-0.00	0.10	0.01	0.15	1		
org_num	1.98	0.97	-0.06	-0.04	0.06	0.01	0.03	0.06	0.10	0.32	0.06	1	
org_rank	0.29	0.46	0.05	0.02	0.01	-0.01	-0.01	0.09	0.01	-0.17	-0.03	-0.10	1

表3 基础回归结果($N=669$)

Tab.3 Basic regression results ($N=669$)

变量	单变量	时间一审稿人	控制变量
	$m1$	$m2$	$m3$
ncon_num	-0.082** (0.037)	-0.084** (0.036)	-0.091*** (0.032)
ref			0.019*** (0.003)
page			0.017 (0.022)
au_num			-0.095*** (0.031)
fund			0.071 (0.088)
org_num			-0.056 (0.037)
org_rank			-0.024 (0.090)
cons	2.289*** (0.047)	2.112*** (0.133)	1.312*** (0.263)
lnalpha	-0.106 (0.065)	-0.138** (0.067)	-0.321*** (0.080)
loglikelihood	-2232.64	-2222.45	-2164.25
aic	4471.29	4462.90	4358.50
bic	4484.80	4503.45	4426.09
year	No	Yes	Yes
re_num	No	Yes	Yes

注: **表示 $p < 0.05$, ***表示 $p < 0.01$ 。下同。

4 结束语

4.1 研究结论

本文基于《心理学报》的审稿意见数据,设计了非

低的AIC值意味着更优的模型”的规则,负二项模型是更好的选择,而且Poisson回归结果仍显示非共识度和被引量呈负向关系; $m5$ 给出了OLS估计结果,结果同样成立。其次,在 $m6$ 中更换了自变量,以非共识度(长度)替代非共识度(数量),系数的大小和显著性未发生明显变化。 $m7$ 更换了因变量,以累积被引量替代了两年被引量,系数仍然为负且显著。最后,为了消除每篇论文的审稿意见总量的影响,本文在 $m8$ 和 $m9$ 中进一步考虑了每篇论文的意见总数量(t_num)和意见总长度(t_len),由于审稿人数量与意见总数量和总长度所反映的信息高度一致,在 $m8$ 和 $m9$ 中删除了审稿人数量。结果显示,在加入去中心化后的总数量和总长度后,审稿意见非共识度(数量)与被引频次间的负向关系仍在至少0.05水平上显著。表4的结果表明,在经过一系列稳健性检验后,结果仍然与表3一致,假设进一步得到验证。

共识度指标,采用负二项回归分析探讨了非共识与论文影响力的关系。在经过描述性统计和回归分析后,发现非共识度和论文影响力间存在显著的相关关系。具体而言,论文的非共识度越高,则其被引相对更低。此外,本文发现《心理学报》审稿人的审稿意见的数量和长度的中位数为7条和710字,随着审稿轮数的递增,审稿人的平均意见数量和长度都存在下降趋势。就审稿人而言,《心理学报》首轮审稿中的审稿人主要以2~3人为主,随着审稿人数量的增加,非共识度也在增加。

本文的学术贡献体现在以下几个方面:首先,在已有文献关注同行评议中非共识现象的基础上,本文重点关注了论文评审中的非共识与论文影响力的关系,揭示了非共识的科学学意义,对于同行评审意见挖掘研究有一定的参考价值。其次,本文以中文开放同行评议期刊《心理学报》为样本,并从已有的文献来看,本文是国内较早使用中文开放同行评议文本开展实证研究,对于开放同行评议的开展有着一定的参考意义。本文将基于以上结论,开展

表 4 稳健性检验 (N = 669)
Tab. 4 Robustness tests (N = 669)

变量	Poisson	OLS	非共识度 (长度)	累积被引	意见总数量	意见总长度
	m4	m5	m6	m7	m8	m9
ncon_num	-0.083 ** (0.037)	-0.780 *** (0.275)		-0.096 *** (0.031)	-0.093 ** (0.039)	-0.089 *** (0.033)
ncon_len			-0.072 ** (0.035)			
ref	0.014 *** (0.003)	0.187 *** (0.038)	0.019 *** (0.003)	0.017 *** (0.004)	0.019 *** (0.004)	0.019 *** (0.004)
page	0.028 (0.021)	0.237 (0.234)	0.018 (0.022)	0.034 (0.025)	0.019 (0.024)	0.019 (0.024)
au_num	-0.081 ** (0.034)	-0.814 ** (0.330)	-0.096 *** (0.032)	-0.128 *** (0.036)	-0.093 *** (0.032)	-0.093 *** (0.031)
fund	0.033 (0.104)	0.399 (1.018)	0.070 (0.088)	0.130 (0.094)	0.071 (0.089)	0.070 (0.090)
org_num	-0.062 (0.041)	-0.580 (0.375)	-0.060 (0.037)	-0.044 (0.038)	-0.058 (0.037)	-0.057 (0.037)
org_rank	0.019 (0.095)	0.147 (0.992)	-0.026 (0.091)	-0.034 (0.093)	-0.028 (0.092)	-0.028 (0.091)
t_num					0.009 (0.044)	
t_len						0.006 (0.048)
Cons	1.447 *** (0.252)	0.664 (2.556)	1.315 *** (0.263)	2.557 *** (0.286)	1.384 *** (0.262)	1.385 *** (0.251)
lnalpha			-0.320 *** (0.080)	-0.241 *** (0.092)	-0.307 *** (0.087)	-0.307 *** (0.087)
Adj R ²	—	0.114	—	—	—	—
loglikelihood	-4000.21	-2577.61	-2164.58	-2575.52	-2168.64	-2168.65
aic	8028.42	5183.22	4359.17	5181.05	4365.28	4365.30
bic	8091.50	5246.30	4426.76	5248.64	4428.36	4428.38
year	Yes	Yes	Yes	Yes	Yes	Yes
re_num	Yes	Yes	Yes	Yes	No	No

如下讨论与阐述。

1) 科学的发展与科学规范的适用性。虽然并不能说明非共识的研究即是低质量的研究。实际上,具有正面意见而被录用的文章不一定是高质量文章,被拒稿的文章也不一定是低质量文章^[23],但是稿件拒用率较高的杂志所在的学科往往缺乏对学术标准的共识^[21]。在后学院科学时代,科学知识从学科知识走向公共知识,科学知识生产中的价值负载更加明显^[39]。正是这种价值负载,使得科学知识对价值中立性的寻求更加迫切,从而寻求使科学知识平等地服务于所有人。与此同时,当代科学分工越来越细,造成专家知识的主观性和有限性,进一步造成了同行评议的局限性。来自不同学科的评审专家,即使有着明确的审稿标准,也可能会因主体的不同价值取向而采取相对灵活的标准,这也可能导致非共识的出现^[14]。从科学学的角度审视,科学规范的使用可能会随着领域、专业和学科的知识体系化程度不同而发生变化^[40]。然而,与规范不一致的现象并未否定科学规范的合理性。随着科学的发展,规范化和形式化的表达更有助于学者达成共识。研究者在研究中更加追求规范化和形式化的表达方式,科学理

性也在这一过程进一步得到彰显,科学的普遍性进一步增强^[41]。

2) 期刊应当在同行评议中发挥导引作用。对于期刊而言,期刊可以通过培训,引导审稿人做出更规范的同行评议。审稿人的“非共识”应注明是对创新性或学术贡献的非共识或是其他方面的非共识,且审稿人对文章的创新性程度,创新性在理论或应用上的贡献,在论证或方法上的完善程度,要有着明确的判断。此外,促进审稿人之间的意见交流,也可以在一定程度上减少审稿中的非理性行为,有助于提高审稿质量。对于编辑部来说,当出现非共识现象时,应该依据产生非共识的原因而分别对待^[4]。共识性较高的论文,也需要深入解读,编辑和主编的把关作用应该得到重视。从作者角度来说,作者应在研究中突出对现有知识的边际贡献,有助于审稿专家更好地把握作者的学术思想。开放同行评议 (OPR) 的重要性越发凸显,OPR 过程促进了作者和审稿人的地位平等,有助于克服同行评议中的理解偏差和单向沟通^[42]。与此同时,OPR 也促使审稿人给出更加积极的意见^[43],从而有利于论文质量的提升。

4.2 局限

本文是国内对中文开放同行评议报告的探索性研究,但仍存在一定的局限。第一,尽管当下 OPR 研究往往基于某本期刊或某一会议的录用论文的审稿数据展开^[28,31],但限于数据,本文仅基于《心理学报》一本期刊的数据展开实证研究,而且仅关注被录用的文章,这就有可能限制了研究发现的外推。第二,同行评议的非共识现象更应该基于文本的内容来衡量。然而,正如文中所述,这一点是比较困难和模糊的^[35]。第三,本文主要采用了初审数据,而审稿过程本身是否会对论文质量有所提升方面可能存在瑕疵。尽管同行评议对提升论文质量造成的影响范围可能是有限的,论文的核心创新和研究内容很少通过同行评议得到实质性的改变^[23]。第四,本文的研究结论仅停留在相关性层面,而因果关系的发现则有赖于更加全面的数据。这些都有待在今后的研究中进一步展开讨论。□

参考文献

- [1] 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.
- [2] 李江.发展完善同行评议制度[N].中国社会科学报,2021-04-06(1).
- [3] 张琳,孙梦婷.突破交叉科学研究同行评议困境[N].中国社会科学报,2021-11-23(1).
- [4] 吴述尧.同行评议方法论[M].北京:科学出版社,1996.
- [5] 杨列勋,汪寿阳,席酉民.科学基金遴选非共识研究项目的评估研究[J].科学学研究,2002(2):185-188.
- [6] 刘求实,丁厚德,王玉堂.同行评议中的“非共识”问题研究[J].科技导报,1995(1):40-42.
- [7] 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.
- [8] HOWARD L, WILKINSON G. Peer review and editorial decision-making [J]. British Journal of Psychiatry, 1998, 173: 110-113.
- [9] 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: 1964-1969.
- [10] WADE D, TENNANT A. An audit of the editorial process and peer review in the journal Clinical rehabilitation [J]. Clinical Rehabilitation, 2004, 18(2): 117-124.
- [11] OXMAN A, GUYATT G, SINGER J, et al. Agreement among reviewers of review articles [J]. Journal of Clinical Epidemiology, 1991, 44(1): 91-98.
- [12] SQUAZZONI F, GANDELLI C. Opening the black-box of peer review: an agent-based model of scientist behaviour [J]. Journal of Artificial Societies and Social Simulation, 2013, 16(2): 1-3.
- [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] CHUBIN D E, HACKETT E J. Peerless science: peer review and US science policy [M]. New York: State University of New York Press, 1990.
- [15] American Sociological Review. Granovetter-rejection [EB/OL]. [2021-04-20]. <https://scatter.files.wordpress.com/2014/10/granovetter-rejection.pdf>.
- [16] AKERLOF G. Writing “the market for ‘lemons’”: a personal and interpretive essay [EB/OL]. [2021-04-20]. <https://www.nobelprize.org/prizes/economic-sciences/2001/akerlof/article/>.
- [17] BARBERIS N, HUANG M. Stocks as lotteries: the implications of probability weighting for security prices [J]. American Economic Review, 2008, 98(5): 2066-2100.
- [18] BAILAR J. Reliability, fairness, objectivity and other inappropriate goals in peer review [J]. Behavioral and Brain Sciences, 1991, 14: 137-138.
- [19] LANGFELDT L. The decision-making constraints and processes of grant peer review, and their effects on the review outcome [J]. Social Studies of Science, 2001, 31(6): 820-841.
- [20] HODGSON C. How reliable is peer review? an examination of operating grant proposals simultaneously submitted to two similar peer review systems [J]. Journal of Clinical Epidemiology, 1997, 50(11): 1189-1195.
- [21] 默顿.科学社会学[M].鲁旭东,林聚任,译.北京:商务印书馆,2003.
- [22] MONTGOMERY A A, GRAHAM A, EVANS P H, et al. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference [J]. BMC Health Services Research, 2002, 2(1): 8.
- [23] 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.
- [24] DALTON M S. Refereeing of scholarly works for primary publishing [J]. Annual Review of Information Science and Technology, 1995, 30: 213-250.
- [25] 张光耀,姜春林,王贤文.即时开放获取论文在时间和地理空间上的使用优势分析——以《新英格兰医学期刊》为例[J].中国科技期刊研究,2019,30(8):890-896.
- [26] ZHANG Guangyao, WANG Licheng, XIE Weixi, 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, 2021, ahead-of-print.
- [27] QIN Chenglei, ZHANG Chengzhi. Exploring the distribution regularities of referees’ comments in IMRAD structure of aca-

- demic articles [C] //18th International Conference on Scientometrics and Informetrics Conference, ISSI 2021, July 12, 2021 – July 15, 2021. Leuven, Belgium: International Society for Scientometrics and Informetrics, 2021: 1527-1528.
- [28] 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.
- [29] XU Shenmeng, ZHANG Guangyao, SUN Yao, et al. Understanding the peer review endeavor [J]. Proceedings of the Association for Information Science and Technology, 2019, 56 (1): 316-325.
- [30] BLACK N, VAN ROOYEN S, GODLEE F, et al. What makes a good reviewer and a good review for a general medical journal? [J]. Jama-Journal of the American Medical Association, 1998, 280 (3): 231-233.
- [31] NI Jue, ZHAO Zhenyue, SHAO Yupo, et al. The influence of opening up peer review on the citations of journal articles [J]. Scientometrics, 2021 (126): 9393-9404.
- [32] 李金珍,庄景春,邱炳武.《心理学报》开放性同行评审方式探索及初步成效 [J]. 中国科技期刊研究, 2015, 26 (2): 139-142.
- [33] AZAR O H. The slowdown in first-response times of economics journals: can it be beneficial? [J]. Economic Inquiry, 2007, 45 (1): 179-187.
- [34] 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.
- [35] 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.
- [36] ZHANG Lin, SHANG Yuanyuan, HUANG Ying, et al. Gender differences among active reviewers: an investigation based on publons [J]. Scientometrics, 2022, 127: 145-179.
- [37] 郑兴东,陆伟,贺宪民,等. 基金项目同行评议中专家非共识性的度量研究 [J]. 中华医学科研管理杂志, 2005 (3): 141-144.
- [38] 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.
- [39] JASANOFF S. States of knowledge [M]. London: Routledge, 2004.
- [40] MITROFF I. Norms and counter-norms in a select group of the Apollo moon scientists: a case study of the ambivalence of scientists [J]. American Sociological Review, 1974, 39 (4): 579-595.
- [41] 桑明旭. 科学精神的谱系: 默顿、齐曼与马克思——兼论科学的公共性价值追求 [J]. 科学经济社会, 2014, 32 (3): 32-37.
- [42] 杜杏叶,李贺,王玲,等. 中国学者对学术论文公开同行评议的接受度研究 [J]. 图书情报工作, 2018, 62 (2): 73-81.
- [43] MATSUI A, CHEN E, WANG Y, et al. The impact of peer review on the contribution potential of scientific papers [J]. Peer J, 2021, 9: e11999.
- 作者简介: 张光耀,男,1995年生,博士生。研究方向: 同行评议。谢维熙,女,1999年生,博士生。研究方向: 科学计量学。夏鑫璐,女,1999年生,硕士生。研究方向: 科学计量学。耿屿,男,1992年生,博士生。研究方向: 计算社会科学。姜春林,1970年生,教授。研究方向: 科技评价。王贤文(通信作者, Email: xianwenwang@dlut.edu.cn),男,1982年生,博士,教授,教育部青年长江学者。研究方向: 科学计量学。
- 录用日期: 2022-03-18
- (上接第93页)
- [35] 因环境监测数据造假山西临汾市长再被环境部约谈 [EB/OL]. [2018-08-06]. <https://baijiahao.baidu.com/s?id=1608020380504901888&wfr=spider&for=pc>.
- [36] 360 天眼实验室用大数据保障数据驱动安全 [EB/OL]. [2015-07-01]. <http://scitech.people.com.cn/n/2015/0701/c1057-27233391.html>.
- [37] 姜策群,周承聪. 信息服务机构信息生态位的优化策略 [J]. 情报理论与实践, 2011, 34 (6): 1-3, 7.
- [38] 顶级校企合作面临三大挑战 [EB/OL]. [2021-05-17]. <https://baijiahao.baidu.com/s?id=1699999719001641404&wfr=spider&for=pc>.
- [39] The Royal Society. Science as an open enterprise [EB/OL]. [2020-01-08]. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>.
- [40] Science 关注: 巨额科研经费使用不当,著名科学家被开除 [EB/OL]. [2020-12-04]. <http://med.china.com.cn/content/pid/220367/tid/1026>.
- [41] 制裁华为打压华为美国主流媒体都看不下去了 [EB/OL]. [2019-05-20]. <http://gd.sina.com.cn/sztech-news/hlw/2019-05-20/detail-ihvnews3130349.shtml>.
- [42] 国家基础科学数据共享服务平台 [EB/OL]. [2018-04-19]. http://www.ecas.cas.cn/wxjj/yxxxt/201804/t20180419_4539187.html.
- 作者简介: 孙晓彬,女,1998年生,硕士生。研究方向: 科研大数据。佟泽华(通信作者, Email: tzhch@163.com),男,1975年生,博士,教授,博士生导师。研究方向: 情报分析,科研大数据。韩春花,女,1979年生,硕士,副教授。研究方向: 企业管理,情报分析,科研大数据。石江瀚,男,1997年生,硕士生。研究方向: 科研大数据。
- 录用日期: 2022-05-06