

科学评价视角下 F1000、Mendeley 与传统文献计量指标的比较*

宋丽萍 王建芳 王树义

摘要 为了探讨同行评议、影响计量学以及传统文献计量指标在科学评价中的有效性,本文选取 F1000、Mendeley 以及 Web of Science、Google Scholar 数据库,采用 SPSS 19.0 软件,将心理学与生态学的 1,033 篇论文的同行评议结果即 F1000 因子、Mendeley 阅读统计、期刊影响因子,以及 Web of Science、Google Scholar 数据库中被引频次进行相关分析。结果表明:同行评议结果、传统引文分析指标以及以 Mendeley 为代表的影响计量指标具有低度正相关性,这意味着上述指标在科学评价中审视角度的不同以及数字时代科学评价的多维构成;心理学筛选数据中 F1000 因子与期刊影响因子相关度几近为 0,这一结论进一步证实了期刊影响因子与单篇论文影响力的严重背离;生态学与心理学指标相关分析结果的不同折射出科学评价中自然科学、社会科学的差异。图 3。表 4。参考文献 10。

关键词 科学评价 F1000 Mendeley 期刊影响因子 WoS Google Scholar

分类号 G250

Observation of F1000, Mendeley and Traditional Bibliometric Indicators from the Perspective of Scientific Assessment

Song Liping, Wang Jianfang & Wang Shuyi

ABSTRACT To investigate the usefulness of some measures for post-publication evaluations, the authors choose 1,033 ecology and psychology papers in F1000, and compare the F1000 quality filtering results and Mendeley usage data with traditional bibliometric indicators. The finding shows that there are lower positive associations among the initial assessment of F1000 article factor by expert assessment and the citation counts from WoS and Google Scholar and Mendeley user counts, which suggests that these indicators measure different perspectives of research. It also turns out the small correlations between F1000 article factor and journal impact factor in psychology, thus confirms that journal impact factors are not suitable to represent individual article impact. The result also indicates that there are significant differences between natural science and social science in the point of scientific evaluation. 3 figs. 4 tabs. 10 refs.

KEY WORDS Scientific evaluation. F1000. Mendeley. Journal impact factor. WoS. Google Scholar.

长期以来,期刊影响因子(Journal Impact Factor, JIF)一直被视作科学评价的重要标准。然而伴随着数字时代的到来,英国从同行评议

为基础的 Research Assessment Exercise(RAE)到以计量为主体的 Research Excellence Framework(REF)的变革,以及美国前总统科技顾问约翰·

* 本文系国家社会科学基金项目“科技政策视角下科学评价指标及方法研究”(编号:12BTQ033)的研究成果之一。

通讯作者:宋丽萍,Email:slp-lp@163.com

马伯格以计量为基础的第二代科技政策的提出^[1],说明 JIF 评价标准的权威性受到质疑。2012 年,美国信息科学技术学会会刊(JASIST)刊登加拿大魁北克大学蒙特利尔分校 Lozano 等人的文章,文中以 Web of Science(WoS)1900—2011 年间的 819,369,970 篇论文为样本,就论文的被引次数与 JIF 的相关性展开分析,指出:1902 至 1990 年,二者相关性渐强;而进入 20 世纪 90 年代,二者相关性变弱^[2]。2013 年 2 月,英国高等教育基金委员会 REF 项目主管声明,2014 年 REF 实施中将不再采用 JIF^[3]。凡此种显示,将 JIF 等同于论文及其作者研究水平的做法已然遭到学术界的强烈质疑。因此,本文以 F1000、Mendeley 为例,结合近期论文层面科学评价(Article-level metrics)的发展,探索同行评议、影响计量指标(鉴于新的指标不断出现,本文采用鲁索博士 influmetrics 一词,而非 Altmetrics)^[4]以及传统文献计量指标在科学评价中的有效性及相关性,以期对数字时代的科学评价提供借鉴。

1 相关背景

Mendeley,基于云的文献管理工具,同时也称学术社交网络平台。其名称源于两位科学家,即俄国化学家门捷列夫(Mendeleev)和奥地利遗传学家孟德尔(Mendel)。2008 年 8 月 Mendeley 的 beta 版发布,2012 年 8 月其机构版发布,2013 年为出版界巨头爱思威尔(Elsevier)公司收购。其中,单篇论文阅读统计(reader counts)实时反映了文章在科学社区内的学术影响力,填补了自论文发表之日起至 WoS 论文被引信息公布之间的空白。

F1000,创新型生物医学文献学术评价体系,系英国现代生物出版集团(BioMed Central)出版,以事后评议方式(Post Publication Peer Review,PPPR)为研究人员甄选生物学与医学领域中重要的论文。该数据库采用加权平均的方法将专家意见汇总,计算出每篇论文的 F1000 因子(F1000 Article Factor,FFa)作为论文同行评议结果,通过信息的二次过滤指引人们快速

检索、发现该领域内有重要价值和前沿性的文献。

本文以 F1000 与 Mendeley 分别作为同行评议与影响计量学的代表,同时选取 JIF 以及 WoS、Google Scholar(GS)中被引频次作为传统文献计量指标的典型加以比较。

2 数据的采集与统计

文中选取 F1000 数据库中的生态学、心理学文献为统计对象,回溯年限为 2007 年,以兼顾自然科学和社会科学引用行为的差异并保证 5 年的被引区间统计。为此,通过 F1000 高级检索界面,设定检索式为:article publication data = 2007 and subject = "psychology",article publication data = 2007 and subject = "ecology"。检索结果为生态学文献 732 篇,心理学文献 301 篇,共计文献 1,033 篇。对于上述记录,首先采用计算机编程语言 Java 抓取了文献的篇名、期刊名、FFa、作者等信息,并将上述信息导入 Excel,利用第三方软件包 org.apache 实现对 Word 与 Excel 的读写,继而利用 2013 年 6 月发布的 Publish or Perish 4.1.2 版(POP),借助其中期刊引用分析功能(journal impact analysis)通过刊名、篇名检索每篇文章的被引量,检索日期为 2013 年 6 月,鉴于 POP 输出结果 1,000 条的上限,部分文献以篇名在 GS 中直接获取被引频次^[5]。同样以篇名为检索字段,利用 Mendeley 数据库获取每篇文献的读者数量,在 WoS 数据库中检索每篇论文 2007 年至 2013 年 6 月间的总被引频次,对于相同主题的文章,通过作者、发表年限、刊名等信息进一步加以区分与确认。此外,将表格中的期刊汇总,借助期刊引用报告(Journal Citation Reports,JCR)逐条查询获取了对应期刊的两年影响因子。

在上述统计基础上,本文主要探讨如下问题:

①FFa、Mendeley 阅读统计、JIF、被引频次在科学评价中的有效性及相关性;

②以生态学 and 心理学为代表的自然科学和

社会科学在科学评价方面的异同;

③相关系数反映的同行评议、影响计量指标、传统引文分析指标科学评价维度的异同,据此就科学评价的构成要素做出分析。

对数据进行正态性分析(从略),分析结果表明显著性水平小于 0.05,因此无论是 FFa 还是 JIF、被引频次、Mendeley 阅读统计均不符合正态分布,应选取非参数的斯皮尔曼检验为宜。

3 数据的处理

依据上述统计结果,借助 SPSS 19.0 软件,对两个学科数据进行分析。首先采用 K-S 检验

3.1 两学科指标相关性分析

表 1、表 2 分别为生态学和心理学文献斯皮尔曼检验结果。

表 1 生态学文献指标相关分析结果

			FFa	JIF	Mendeley	WoS	GS
Spearman 相关系数	FFa	相关系数	1.000	.346 **	.257 **	.336 **	.345 **
		Sig. (双侧)	.	.000	.000	.000	.000
		N	728	728	716	725	726
	JIF	相关系数	.346 **	1.000	.448 **	.561 **	.537 **
		Sig. (双侧)	.000	.	.000	.000	.000
		N	728	728	716	725	726
	Mendeley	相关系数	.257 **	.448 **	1.000	.719 **	.725 **
		Sig. (双侧)	.000	.000	.	.000	.000
		N	716	716	716	714	714
	WoS	相关系数	.336 **	.561 **	.719 **	1.000	.955 **
		Sig. (双侧)	.000	.000	.000	.	.000
		N	725	725	714	725	723
	GS	相关系数	.345 **	.537 **	.725 **	.955 **	1.000
		Sig. (双侧)	.000	.000	.000	.000	.
		N	726	726	714	723	726

注:**表示在置信度(双侧)为 0.01 时,相关性是显著的。

表 2 心理学文献指标相关分析结果

			FFa	JIF	Mendeley	WoS	GS
Spearman 相关系数	FFa	相关系数	1.000	.267 **	.278 **	.274 **	.285 **
		Sig. (双侧)	.	.000	.000	.000	.000
		N	297	297	293	297	297
	JIF	相关系数	.267 **	1.000	.550 **	.640 **	.605 **
		Sig. (双侧)	.000	.	.000	.000	.000
		N	297	297	293	297	297

续表

			FFa	JIF	Mendeley	WoS	GS
Spearman 相关系数	Mendeley	相关系数	.278 **	.550 **	1.000	.716 **	.710 **
		Sig. (双侧)	.000	.000	.	.000	.000
		N	293	293	293	293	293
	WoS	相关系数	.274 **	.640 **	.716 **	1.000	.958 **
		Sig. (双侧)	.000	.000	.000	.	.000
		N	297	297	293	297	297
	GS	相关系数	.285 **	.605 **	.710 **	.958 **	1.000
		Sig. (双侧)	.000	.000	.000	.000	.
		N	297	297	293	297	297

注：**表示在置信度(双侧)为 0.01 时,相关性是显著的。

3.2 FFa 为 3 以上的论文指标相关分析

所得数据集对应指标相关分析结果如表 3、表 4

以 FFa ≥ 3 为阈值对上述样本进行筛选, 获得生态学文献 147 篇, 心理学文献 51 篇。筛选

所示:

表 3 147 篇生态学文献指标相关性分析结果

			FFa	JIF	Mendeley	WoS	GS
Spearman 相关系数	FFa	相关系数	1.000	.271 **	.306 **	.338 **	.355 **
		Sig. (双侧)	.	.001	.000	.000	.000
		N	147	147	146	147	147
	JIF	相关系数	.271 **	1.000	.221 **	.351 **	.351 **
		Sig. (双侧)	.001	.	.007	.000	.000
		N	147	147	146	147	147
	Mendeley	相关系数	.306 **	.221 **	1.000	.581 **	.617 **
		Sig. (双侧)	.000	.007	.	.000	.000
		N	146	146	146	146	146
	WoS	相关系数	.338 **	.351 **	.581 **	1.000	.933 **
		Sig. (双侧)	.000	.000	.000	.	.000
		N	147	147	146	147	147
	GS	相关系数	.355 **	.351 **	.617 **	.933 **	1.000
		Sig. (双侧)	.000	.000	.000	.000	.
		N	147	147	146	147	147

注：**表示在置信度(双侧)为 0.01 时,相关性是显著的。

表 4 51 篇心理学文献指标相关性分析结果

			FFa	JIF	Mendeley	WoS	GS
Spearman 相关系数	FFa	相关系数	1.000	.080	.388 **	.362 **	.292 *
		Sig. (双侧)	.	.578	.005	.009	.037
		N	51	51	50	51	51
	JIF	相关系数	.080	1.000	.360 *	.478 **	.482 **
		Sig. (双侧)	.578	.	.010	.000	.000
		N	51	51	50	51	51
	Mendeley	相关系数	.388 **	.360 *	1.000	.509 **	.498 **
		Sig. (双侧)	.005	.010	.	.000	.000
		N	50	50	50	50	50
	WoS	相关系数	.362 **	.478 **	.509 **	1.000	.959 **
		Sig. (双侧)	.009	.000	.000	.	.000
		N	51	51	50	51	51
	GS	相关系数	.292 *	.482 **	.498 **	.959 **	1.000
		Sig. (双侧)	.037	.000	.000	.000	.
		N	51	51	50	51	51

注: ** 表示在置信度(双测)为 0.01 时,相关性是显著的; * 表示在置信度(双测)为 0.05 时,相关性是显著的。

4 结果与讨论

首先,从总体上看,FFa、Mendeley 阅读统计、JIF、被引频次相关系数均在 0 和 1 之间,即上述变量两两之间存在正相关关系。如果说同行评议与文献计量作为科学评价的两种主要方法已成为学术界的共识,那么这一结果为以 Mendeley 为代表的新成员在科学评价中的有效性提供了依据。进而就相关系数数值而言,生态学中 FFa 与 WoS 被引频次、GS 被引频次以及 JIF 之间均属 0.3 以上的低度相关,可以说具有一定的关联度;虽然心理学中 FFa 与 JIF、Mendeley 读者数量、WoS 被引频次、GS 被引频次四指标显著性水平平均小于 0.05,但相关系数都在 0.3 以下,说明相关度极弱,从统计学上可以视为不相关。指标在总体变化方向相同前提下的低度相关或者不相关,标志着同行评议、被引频次与影响计量在科学评价中审视视角的不同,从而标志着数字时代科学评价的多维构成。

其次,从被引指标来看,表 1、表 2 相关分析

体现出生态学、心理学 WoS 与 GS 中被引频次的高度相关,两学科中二者相关系数分别为 0.955、0.958, WoS 具有权威性与高度的选择性,而 GS 以施引源广泛著称。图 1 以生态学为例说明了二指标的和諧与一致。这样看来,澳大利亚生态学教授 Vanclay 建议今后澳大利亚高等教育机构科研能力测评系统——ERA 实施中以 GS 作为评价依据也是事出有因^[6]。这就为 GS 在科学评价中的应用提供了有力依据,更为重要的是,这说明 WoS 与 GS 从同一维度反映了论文层面的学术影响力^[7]。

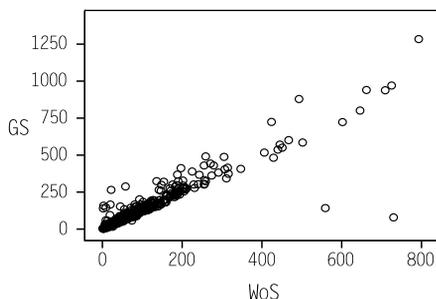


图 1 生态学 WoS 与 GS 相关性

第三,就 JIF 而言,表 1、表 2 体现了 JIF 与论文被引量、Mendeley 读者统计的中度相关。与表 1、表 2 相比,表 3、表 4 筛选数据集分析结果最为显著的变化即是 JIF 与 FFa 的关系,生态学从 0.346 的低度相关落入 0.271 的不相关,心理学则从 0.267 下降为 0.08,几近为 0,图 2 清晰地体现了心理学中二指标的严重背离。具体地说,如果以同行评议为准绳,JIF 将与专家对于论文质量的判断大相径庭,这就意味着,以 JIF 等同于论文影响力将导致单篇论文质量信息的部分或完全失真,这一现象在以心理学为代表的社会学中尤为明显。那么,学术界一直以来墨守的 JIF 等同于论文质量的成规将被打破。该结论再一次印证了文献计量学家 Raan 的观点,即“如果存在一个文献计量学家的共识,那就是永远不要用刊物的影响因子来评价一篇论文或某个研究人员的学术表现”^[8]。进一步来说,随着数字时代论文级学术影响力挖掘能力的增强,JIF 与论文质量及其影响力的不等价性趋于明显。那么,或许在不远的将来,JIF 将完成其作为论文及其作者研究水平替代品的使命,逐渐退出历史舞台。

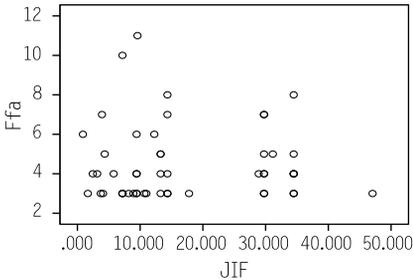


图 2 心理学 51 篇文献 FFa 与 JIF 的关系

第四,就新型影响计量指标而言,两个学科中,Mendeley 阅读统计与被引量以及 JIF 皆呈中度相关(见图 3)。这一结果在确认社交媒体影响指标的科学评价成员身份的同时,亦揭示了 Mendeley 阅读统计等指标与基于引用的传统文献计量指标的差异性。首先,Mendeley 读者统计与被引频次的中度相关说明一定程度上 Mendeley 能够预测论文后期引用的情况。然而,差异也是客观存在的:其一,作为 Web2.0 的产物,

Mendeley、CiteUlike、Twitter 等学术社交网络工具实时反映了论文在更广泛范围内的影响力;其二,与 JIF 有所不同,单篇研究论文成为 Mendeley、Zotero、CiteUlike 系统中的主角,因此使论文独立于其载体之外,从而彰显了论文自身的影响力;其三,Mendeley 等社交媒体影响指标与传统文献计量指标代表了信息在科学共同体中不同的扩散模型。根据美国科学史学家普莱斯的引用峰值理论:文章发表后两年被引用的次数最多,然后逐渐减少,进入半衰期、老化期^[9]。然而基于 PLOS 旗下六种期刊的网络利用统计表明:论文发表第一个月即达到利用峰值,第二个月则显示出明显的利用衰变^[10]。由此可见,被引峰值与利用峰值在论文问世后不同的时间段出现。综上,Mendeley 作为科学评价的一个维度反映了论文当前以及潜在的学术影响力。

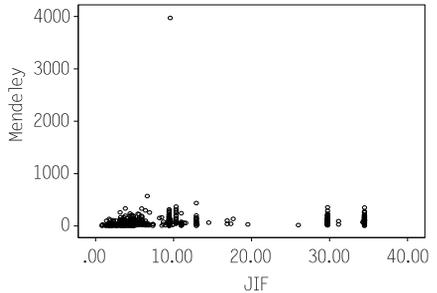


图 3 生态学中 Mendeley 读者统计与 JIF 的关系

第五,表 1、表 2 生态学与心理学指标相关分析结果表明,生态学中 FFa 与 JIF、被引频次相关系数明显高于心理学,这一差异折射出自然科学、社会科学在科学评价中的不同。究其根源,盖因二者研究对象与研究范式不同所致,即普莱斯所谓:“自然科学与人文社会科学不同实质内容确立了不同的信息集合、不同的社会交换工具”,因此表明社会科学评价不能完全套用自然科学评价指标与标准^[7]。事实上,欧美国家已认识到二者在科学评价中的差异性,英国在基于计量的 REF 高等教育评价体系实施过程中,对于自然科学采用文献计量法,而在艺术、人文、社会科学等领域则采用同行评议为主、引文分析为

辅的方式。那么,生态学与心理学中指标相关度的不同将意味着自然科学与社会科学中同行评议、文献计量、影响计量三方配比的不同。

5 结论

通过以上分析,本文得出四点结论。

①从统计学视角,同行评议结果与传统引用及影响计量指标具有正向相关性,表明上述指标在科学评价中的有效性与可行性;同时低度相关的结果表明三者从不同维度描述了论文层面的学术影响力,表明数字时代科学评价的多维格局。

②单篇论文同行评议结果、影响计量指标及被引指标可获性的增强,标志着 Web2.0 环境

下论文层面科学评价的逐渐形成与完善,这将导致 JIF 在科学评价中地位的弱化。伴随着 JIF 主角地位的动摇,科学评价将逐步进入论文层面评价的时代。

③就自然科学和社会科学而言,评价指标相关性的差异蕴含着自然科学和社会科学评价方式的不同,因此同行评议、文献计量、影响计量三方配比应随学科不同而变化。

④从总体上看,科学评价的多维构成尚属研究中的问题。对此,也许“盲人摸象”是最贴切的比喻:新的指标在不断出现,新的视角在不断形成,但是每个人、每个指标所能观察到的仅仅是科学评价的一个侧面。然而有一点是值得肯定的,随着论文级计量影响力的扩大,我们将越来越接近科学评价的本来面目。

参考文献

- [1] Kaye H F, Julia I L, John H M, et al. The science of science policy: a federal research roadmap[EB/OL]. [2014-06-10]. http://scienceofsciencepolicy.net/sites/all/themes/sosp_theme3/userfiles/SoSP_Roadmap.pdf.
- [2] George A L. The weakening relationship between the impact factor and papers' citations in the digital age[J]. Journal of the American Society for Information Science and Technology, 2012, 63(11): 2140-2145.
- [3] Impact factors, research assessment and an alternative to REF 2014[EB/OL]. [2013-06-30]. <http://blogs.lse.ac.uk/impactofsocialsciences/2013/02/08/impact-factors-and-an-alternative-to-ref-2014/>.
- [4] Ronald R, Fred Y Ye. A multi-metric approach for research evaluation[J]. Chinese Science Bulletin, 2013, 58(26): 3288-3290.
- [5] Publish or Perish. [2013-07-18]. <http://www.harzing.com>.
- [6] Peter J. Using Google Scholar for journal impact factors and the h-index in nationwide publishing assessments in academia-Siren songs and air-raid sirens[J]. Online Information Review, 2012, 36(3): 462-478.
- [7] Li Xuemei, Thelwall M. F1000, Mendeley and traditional bibliometric indicators[C]//Proceedings of the 17th International Conference on Science and Technology Indicators, 2012: 451-551.
- [8] Van Noorden R. A profusion of measures[J]. Nature, 2010, 465(7300): 864-866.
- [9] 邱均平. 中国学术期刊评价研究报告[M]. 北京: 科学出版社, 2011: 59-62. (Qiu Junping. A report on Chinese academic journals evaluation[M]. Beijing: Science Press, 2011: 59-62.)
- [10] Yan K K, Gerstein M. The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics[J/OL]. Plos One, 2011, 6(5): e19917. [2014-06-10]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3095621/>.

宋丽萍 天津师范大学管理学院教授。通讯地址:天津师范大学管理学院。邮编:300387。

王建芳 中国科学院国家科学图书馆副研究馆员。

通讯地址:北京市中关村北四环西路33号。邮编:100080。

王树义 天津师范大学管理学院讲师。通讯地址:天津师范大学管理学院。邮编:300387。

(收稿日期:2013-11-15)