

2021 年世界科技评估研究前沿观察

科技部科技评估中心编写组

(一) 概述

2021 年,《研究评估》共刊载了 43 篇学术文章(个别非研究类文章未纳入本报告)。为呈现科技评估领域的研究热点和重要机构,编写组对《研究评估》刊载文章的关键词、作者、国别、机构等信息进行了统计。结果显示,2021 年《研究评估》期刊发文涉及 24 个国家、92 家机构、188 个关键词和 147 位作者。发文作者涉及西班牙、美国、德国、英国、荷兰、土耳其、越南等 24 个国家,各国作者发文数量见图 1。中国方面,来自浙江理工大学、武汉大学等机构的作者共发表了 2 篇论文。

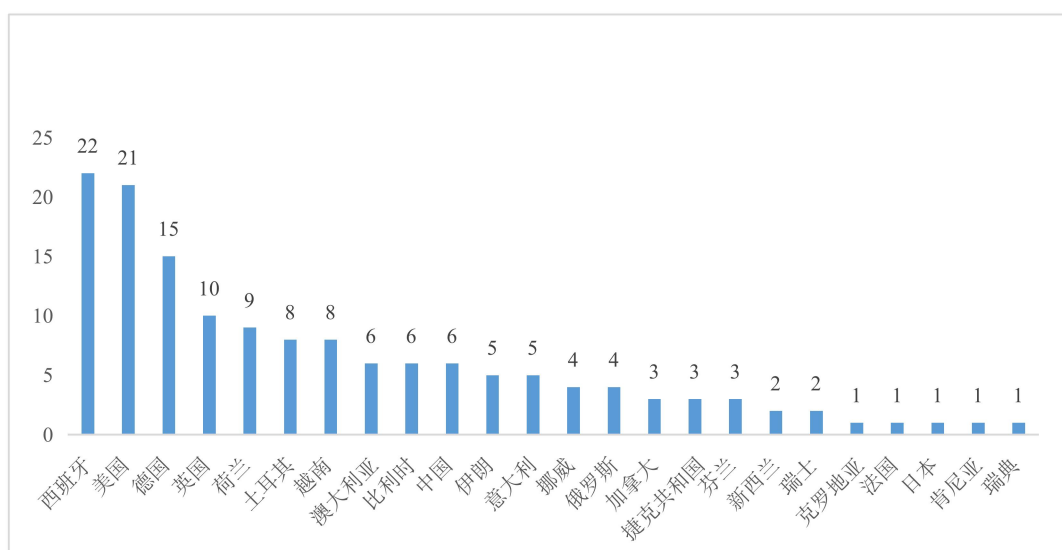


图 1 《研究评估》期刊发文较多的国家(2021)

2021 年《研究评估》发文作者来自全球 92 家机构,包括大学、研究所等类型。统计分析发现,来自美国华盛顿大学(University of Washington)的研究人员活跃度最高,共有 7 人发文;其次是美国国际开发署(The US Agency for International Development),有 6 人发表文章。美国弗雷德·哈钦森癌症研究中心(Fred Hutchinson Cancer

Research Center)、西班牙巴塞罗那国立大学附属医院(Hospital Clínic-Universitat de Barcelona)、伊朗医科大学(Iran University of Medical Sciences)、土耳其中东技术大学(Middle East Technical University)、西班牙杜斯托大学(University of Deusto)、澳大利亚悉尼大学(University of Sydney)等机构的作者也较为活跃,参与文章发表的人数为5人。具体排序情况见图2。

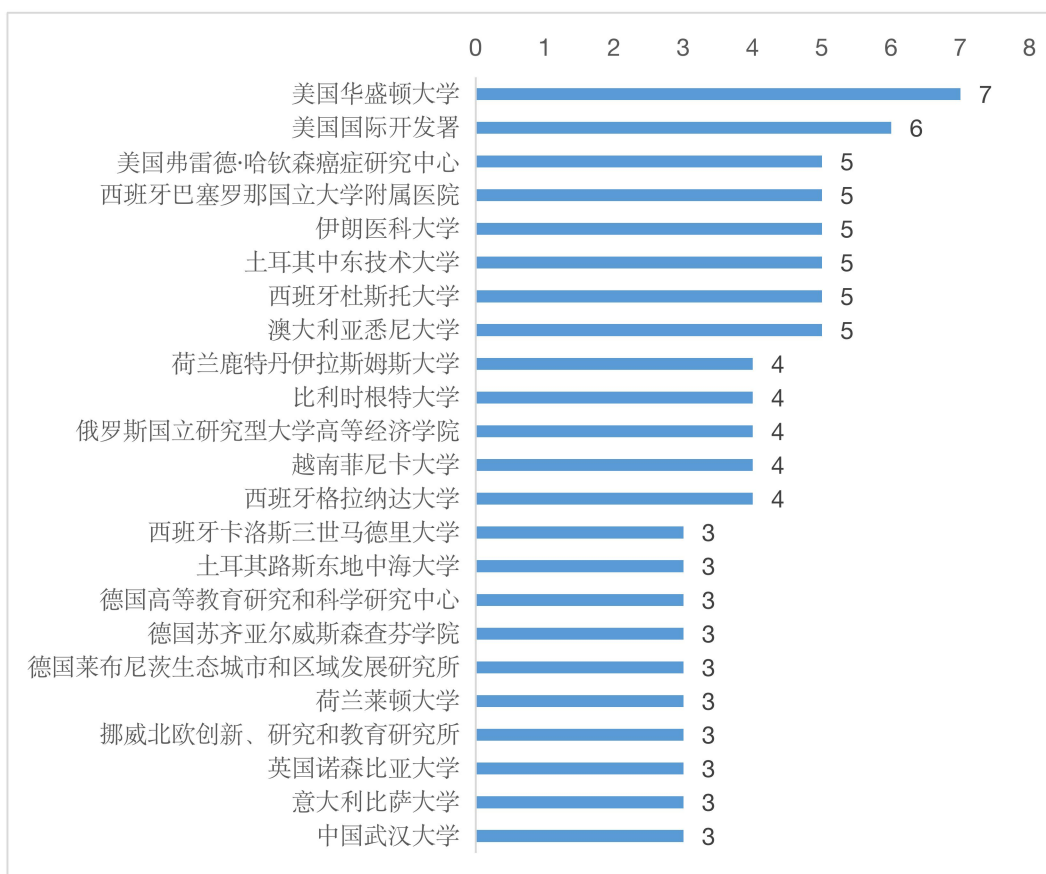


图2 在《研究评估》发文作者数量较多的机构 (2021)

备注: 图中数字代表该机构发表文章的作者人数

2021年发表文章中出现的关键词共计188个,其中17个关键词出现2次以上(包括2次)。出现频次较高的关键词包括研究评价、社会影响、同行评议等(见表1)。

表1 《研究评估》出现次数较多的关键词 (2021)

关键词	中文	出现频次
research evaluation	研究评估	7
societal impact	社会影响	6
peer review	同行评议	5
higher education	高等教育	4
impact	影响	4
impact assessment	影响评估	4
research impact	研究影响	3
social sciences and humanities	社会科学及人文	3
collaborative research	合作研究	2
evaluation	评估	2
evidence-based policy	循证政策	2
interdisciplinary research	交叉学科研究	2
public health	公共卫生	2
research productivity	研究生产力	2
responsible research and innovation	负责任的研究和创新	2
transdisciplinary research	跨学科研究	2
university	大学	2

编写组进一步对《研究评估》刊载文章进行主题聚类，发现 2021 年的刊载文章突出 3 个主题，即跨学科研究评估、影响评估和同行评议。以下围绕这 3 个主题分别推荐若干代表性文章，介绍文章摘要，读者可查阅全文深入了解有关内容。

(二) 跨学科研究评估

Aihua Chen and Xiaoting Wang 发表《促进跨学科合作对高校科研团队科研生产力的影响:政府援助的调节作用》一文。文章研究了促进跨学科合作与研究生产力之间的联系,以及政府支持在这一联系

中的作用。文章以 314 名中国高校科研团队成员为研究对象，发现管理、科研活动、评估、团队建设以及与产业界共同促进跨学科合作均对高校科研团队的科研生产力有显著的正向影响。其中，评估活动对科研生产力的影响最为显著。研究对于政府和大学促进跨学科合作具有启示。

Seokbeom Kwon 等发表了《跨学科知识组合和新兴技术主题：对减少研究评估中的不确定性的启示》一文。文章提出了新兴技术主题的新指标，并测试其与跨学科和非典型研究组合的关系。研究通过 3 个不同跨学科领域的科学出版物元数据来进行测试，分别是纳米药物输送、合成生物学和自动驾驶汽车。进一步研究发现，跨学科和非典型知识组合对新兴技术主题的贡献随着时间的推移而增加或保持不变。

Zhanglin 等发表了《关于跨学科和影响力之间的关系：对学术影响和更广泛影响的不同效应》。跨学科研究（IDR）越来越被认为是学术兴趣和国家科学政策的核心。研究分析了跨学科对文章层面的引文影响（尤其是 WoS 引文）和更广泛的影响（尤其是 PLoS 使用）的影响。从多样性、平衡性和差异性 3 个不同元素以及整体的综合多样性来衡量 IDR。带有领域固定效应和稳健标准差的负二项回归分析结果表明，跨学科研究对学术影响和更广泛影响都有积极作用。从一段时间的趋势分析结果来看，较高的跨学科出版物往往能吸引更多的引文，并具有更高的公共科学图书馆（PLoS）使用率。

Martina Schäfer 等发表了《跨学科研究社会影响的系统分析》。文章介绍了一种启发式方法，该方法可作为工具来研究事前和整个研究过程中预期效果。启发式方法包括发生影响（一级、二级和三级）的时间和空间维度，并描述了可能的影响形式。该实证研究涉及不同主题领域的 16 个已完成的跨学科研究项目。研究建议对常用类别进行区分，如“学习影响”，并将二级、三级影响操作化，以确保跨学科

研究团队使用术语更加统一。

(三) 影响评估

Brian M Belcher 等发表了《理解和评估以问题为导向的综合研究计划的影响：概念和考虑》一文。研究提出了一种多方面、多尺度、基于理论的评估方法。其包括嵌套的项目层面和计划层面的变革理论 (ToC)；研究质量评估；基于理论的产出评估，用于实证检验变革理论并评估政策、制度和实践影响；FTA (森林、树木和农林业领域) 熟悉的“大 n”创新实验和准实验的影响，以估计大规模潜在影响的事前影响评估，将计划层面的成果与关于发展和保护状况的二手数据合乎逻辑且合理地联系起来。

Andrey Lovakov 等发表《政府对几所顶尖大学的支持是否对高等教育系统产生了更广泛的影响？俄罗斯大学卓越计划评估》文章。研究利用俄罗斯大学卓越计划 (RUEI) 的数据衡量政府重点支持的溢出效应，并证明确实存在更广泛的影响。特别是，研究了那些既不属于 RUEI，也没有得到其直接支持高校的表现。研究还比较了有和没有 RUEI 的各地区大学的表现。在此情况下，我们衡量了 RUEI 在区域层面对高等教育部门的间接影响。研究展示了对出版活动水平的积极影响，这种影响最近已经变得明显。然而，其对年轻教师的比例、发表论文中的国际合作或招生质量没有影响。从更广泛的政策研究/研究评估的角度来看，该研究揭示了卓越计划的系统性效果，这是经常被忽视的。

Jorrit P. Smit 发表了《研究评估中科学和社会价值的产生：社会影响评估方法综述》一文。文章阐述了评估研究的社会影响主要方法的理论方面，特别是它们对研究的科学价值和社会价值之间关系的假设。研究从建构主义的角度分析了 10 种评估研究社会影响的方法。这些方法代表了对知识交流的不同理解，可以用线性、循环和联合生产的模型来理解。此外，评估方法对研究的社会价值使用了多种概念，

这表明与科学价值的不同关系。虽然一些方法依赖于两种类型的价值之间的明确区分，但其他方法，特别是评估调查、ASIRPA、贡献图谱、公共价值图谱和SIAMPI，则考虑到产生社会价值的机制是研究过程的组成部分。研究得出如下结论，评估方法必须在将社会价值划分为一个单独的绩效指标以达到实际目的，和公正对待科学研究中关于科学价值和社会价值的整合之间取得平衡。

Maryam Razmgir 等发表了《探索研究影响模型：一个系统范围的回顾》一文。文章探讨了关于“研究影响”的模型和框架。分析包括1996年至2020年间发表的47项研究和10篇综述。大多数模型是为评价影响的目的而开发的。研究在模型中确定了3种方法，即基于结果的方法、基于过程的方法以及同时利用结果和过程的方法，其中基于结果的方法是影响模型最常用的方法。基于过程的模型主要改编自凯洛格基金会(W.K. Kellogg Foundation)的逻辑模型，具有应用于影响改进的潜能。研究认为，基于过程的方法在提升和加速研究影响方面具有优势，在制定影响模型时考虑这种方法是重要的。

Florence Gignac 等发表了《参与式研究项目的自评估：在线影响评价工具 InSPIRES 的内容验证》一文。InSPIRES H2020 项目开发了一套影响评估工具，包括反映知识民主、公民主导的研究、参与性动力学、变革性改变和完整性等维度的64个条目。本文试图对该工具的内容有效性开展测试，并提出有效性保证建议。小组由9名专家组成，根据以下3个标准对每个条目进行评估：代表性、相关性和清晰度。基于专家评级结果采用 Aiken's V 和 Wilson Score 方法来评估该工具的内容有效性。专家的书面意见文字也一并审查。在小组层面，对照3项标准75%条目被认为是令人满意的。然而，在总体层面，72%条目表明该工具的一些部分是无效的并需要修正。

(四) 同行评议

Axel Philipps 发表了《科学规则，有关科学家对项目资助随机方

法的定性研究》一文。有些学者认为，项目随机资助的评估方式使得随机分配看起来公平、不那么偏袒，并且更支持创新研究。但不知道研究人员对项目随机资助的实际想法，更不知道那些依赖项目资金的科学家的想法。本文探讨了科学家们对“抽签”选择资助的观点，以及他们如何证明自己的支持或反对。他们如何科学地对待某件本身不科学的事情？文章对德国自然科学家进行了以问题为中心的访谈。本文的访谈表明，处于主导地位 and 主导领域地位的科学家或多或少对通过抽签进行选择的想法持开放态度。尽管如此，他们反对纯粹的随机化，因为从他们的角度来看，这与科学原则是不相容的。他们更倾向于将项目随机和同行评议过程相结合，只有在这些条件下，随机资助的项目才能成为科学不可分割和合法的一部分。

Liv Langfeldt 等发表了《指标体系在同行评议中的角色》一文。文章讨论了科学出版物及其引用情况的指标是否是评估工作合理且必不可少的部分。在对 3 个国家进行广泛问卷调查的基础上，分析了研究人员的意见。研究进行了跨学术领域（心脏病学、经济学和物理学）和评估研究内容（确定各个领域的最佳研究成果、评估拨款提案和评估职位候选人）的比较。少数对调查做出回应的研究人员认为这些指标是评估最佳研究项目的依据。尽管如此，在所有研究领域中，绝大多数人表示，在审查拨款方案和评估学术职位候选人时这些指标是很重要或部分重要的。在这些情况下，出版物的引用影响，尤其是出版物的数量是被看重的。这些发现适用于所有分析领域，但经济学家比心脏病学家和物理学家更依赖生产力指标。此外，在文献计量指标上得分较高的审稿人似乎更经常（比其他审稿人）在评估中遵循指标。因此，在计划和使用同行评审时，应意识到评审员，尤其是在指标上得分较高的评审员，在其评估过程中依赖指标，尽管科学界对出版物指标的使用和误用表示担忧。

Serge P.J.M. Horbach 发表了《现在没时间了!Covid-19 大流行期

间稿件同行评审的质量变化》一文。全球性的新冠疫情对科学事业产生了重大影响，包括学术出版和同行评议工作。一些研究评估了这些影响，结果表明与其他研究主题相比，医学期刊对 Covid-19 相关内容大大加快了审查进程。这引起了人们对审查过程的质量和稿件出版标准的质疑和担忧。为了解决这些问题，本研究评估了 2020 年新冠疫情期间发表的与 Covid-19 相关的文章、与 Covid-19 无关的文章以及疫情之前发表的文章的审稿报告和编辑审稿决定通知的质量差异。研究采用了《英国医学杂志》和 eLife 的开放式同行评审模式来研究评审报告、编辑决定、作者回复和读者公开评论的内容。研究发现，在疫情期间或之前发表的与 Covid-19 无关的文章审查过程没有明显差异。然而，Covid-19 和非 Covid-19 相关文章之间确实存在明显差异，包括更少的补充实验要求、更具合作性的意见，以及针对过于强烈的观点的不同建议。总的来说，研究结果表明审稿人和期刊编辑都含蓄或明确地使用不同的质量标准来评估与 Covid-19 有关的稿件，从而改变了对基础研究的主要评估机制，并可能影响其公开传播。

Dietmar Wolfram 等发表了《对发表在公开同行评议期刊上的审稿人评论的探索：审稿语言的特点及评议审查与引文的关系》一文。研究调查了与稿件评审质量评估相关的两个领域。首先，期刊要求审稿人表明身份的政策是否会影响审稿人根据其评论中出现对冲术语¹和研究相关术语的相对频率来评估稿件的优点？第二，评论/审稿人的数量与稿件发表后的研究影响（以引文衡量）之间是否存在关联？研究选取了 2017—2018 年发表在 17 种 OPR 期刊上文章的评论，以考察评论者使用对冲术语和研究相关术语的发生率。结果表明，无论审稿人是否被要求表明自己的身份，对冲术语的相对使用情况并没有区别，这表明审查内容中对冲术语的使用不受期刊要求审查人表明身

¹ “对冲术语”是指使用某些词或短语来表明所呈现的内容被解释为意见、建议或猜想，而不是事实。在讨论公认文献结果时经常使用确定性术语，而在分析新数据时使用对冲术语。

份的影响。在研究相关术语方面观察到存在较大差异。作者将一篇稿件的审稿总数、修订轮次、审稿人数量与该文章自发表以来收到的 Web of Science 引用次数进行了比较。研究结果显示，更多审稿人的审查、进行更多审阅或多轮审阅并不会导致更具影响力的论文。

Marco Seeber 等发表了《评审经验是否会减少研究计划书评估中的分歧？——基于玛丽·斯沃德斯卡·居里行动（MSCA）和 COST 行动项目的分析》一文。研究旨在探讨评审专家本身的哪些特征最能影响评审专家之间的分歧来探究评审分差。研究提出了关于评审专家的经验水平在评估某一特定授权计划的研究计划书时产生影响的假设，即项目评审经验。我们通过研究 2014 年至 2018 年欧盟最重要的两个研究资助项目来检验假设，具体是：地平线 2020“玛丽·斯沃德斯卡·居里行动（MSCA）”3 个资助计划下开展评审的 52488 份研究计划书，以及“欧洲科技行动合作”评审的 1939 份研究计划书。研究发现，（评审专家）对之前某个特定计划有评审经验会显著减少该类计划的评分分歧，而对其他计划中研究计划书的经验，即一般性评审经验，对减少评分分歧没有任何影响。此外，在 MSCA 个人会员中，我们观察到，评审专家在给定征集评审中评估的研究计划书数量与分数分歧之间呈现出倒 U 型关系，当被评的研究计划书数量超过 13 份时，评分分歧显著减少。研究结果表明，在特定计划中的评审经验可以提高评审可靠性，通过微调评审专家的评估分数可以约束无根据的分数分歧。

（五）其他重要主题

Clare McLaren 等发表了《量化技术人员对研究的贡献》一文。文章介绍了一个基于现有标准的简单框架，该框架可以轻松集成到常见的研究工作流程中来获取技术人员的贡献。通过使用 ORCID 标识符、数字对象标识符和贡献者角色分类法，可以量化、识别技术人员的贡献，并为未来研究的规划和管理创建有价值的额外数据。该框架以澳大利亚高等教育体系为背景，以科学、技术、工程和数学为重点，可

以推广到所有学科。

Katharina Hölsche 等发表了《改变科学和社会？转型研究的方法学》一文。在过去几年中，转型研究已成为研究和支撑向可持续性迈进的激进式社会变革的共同视角。研究开发了一个框架，确定了设计标准，并对研究结果、设计和过程进行反思。研究结果为转型研究在实践中的表现以及如何加强转型研究提供了经验和建议：1.重新定义和评估社会影响研究；2.重新设计研究，整合激进社会变革的观点；3.重新增强研究人员和研究伙伴进行社会学习的能力。

Ondřej Dvouletý 等发表了《评估公共补贴对不同规模受支持企业绩效的效果》一文。本文研究了特定公共政策对企业层面绩效的影响，并将企业规模的异质性作为重要决定因素。研究调查了与微型和小型企业相比，大型企业是否从直接财政援助中获益较少。课题组研究了 2007—2013 年间实施的捷克商业和创新运营项目（OPEI）的微观经济效应。研究使用的是一个大型企业级数据集，由 13924 家企业组成，其中 3572 家是受支持的企业（占 OPEI 计划受益企业的 57%）。通过使用有形固定资产及其折旧、销售额、资产回报率和全要素生产率来衡量企业的整体财务绩效。虽然在方案结束两年后，微型企业的所有评价指标都有所改善，但对大型企业来说，补贴的积极作用不大。

以上简要对 2021 年《研究评估》期刊刊载文章的情况进行了介绍，希望通过分析为我国科技评估界同行打开一扇窗户，帮助评估人员拓展国际视野，了解行业动态，跟踪前沿方向。综述仅供评估人员交流学习使用，希望引导评估人员进一步阅研文章全文并有所收获，提高评估和研究水平。