doi:10.3772/j.issn.2095-915x.2026.02.010

国家重点研发计划拟立项与未立项项目同行评议质量比较研究
—— 以农业农村领域为例

开放科学（资源服务）标识码（OSID）

赵婉莹[1] 王立丽[1] 王璐瑶[1] 袁子晗[2,3] 赵勇[4]

1. 中国农村技术开发中心 北京 100045；
2. 中国人民大学信息资源管理学院 北京 100872；
3. 首都师范大学图书馆 北京 100048；
4. 中国农业大学图书馆 北京 100083

摘要：[目的 / 意义] 高质量的同行评议是遴选出高水平科技项目的重要前提，立项阶段的同行评议质量对科研经费资源配置具有重要影响。[方法 / 过程] 基于 2023 年 8 月至 10 月中国农村技术开发中心承接的农业农村领域 8 个国家重点研发计划重点专项 184 个项目评审数据，对比分析了拟立项与未立项项目的同行评议信度。[结果 / 结论] 研究发现，所有项目的单个专家信度系数和整组专家信度系数均较高，拟立项项目的评分一致性普遍高于未立项项目，不同领域的信度系数存在差异。基于此本研究提出进一步明确评审标准、优化专家选择、重点关注异常值较多的领域、分析未立项项目异常值原因以及持续改进评审体系等政策建议，以期推动科技项目同行评议的科学化与规范化发展。

关键词：国家重点研发计划；农业农村领域；同行评议；质量

中图分类号：G35

Comparative Study on Peer Review Reliability of Proposed and Unproposed National Key Research and Development Programs: A Case Study of Agriculture and Rural Areas

ZHAO Wanying[1] WANG Lili[1] WANG Luyao[1] YUAN Zihan[2,3] ZHAO Yong[4]

1. China Rural Technology Development Center, Beijing 100045, China;
2. School of Information Resources Management, Renmin University of China, Beijing 100872, China;

基金项目 中国农村技术开发中心委托项目“农业农村领域科技项目诚信风险防控机制研究”（NCZX2021-001）；中央高校基本科研业务费专项资金资助科研诚信专项（2025TC019）。

作者简介 赵婉莹（1994-），博士，助理研究员，主要研究方向为科技管理、科技政策研究；王立丽（1971-），硕士，主要研究方向为科技管理、科技政策研究；王璐瑶（1994-），博士，主要研究方向为科技政策；袁子晗（1996-），博士研究生，主要研究方向为科研诚信；赵勇（1980-），通信作者，博士，研究馆员，博士生导师，主要研究方向为科研诚信，科技政策，E-mail：zhaoyong@cau.edu.cn。

引用格式 赵婉莹 , 王立丽 , 王璐瑶 , 等 . 国家重点研发计划拟立项与未立项项目同行评议质量比较研究 —— 以农业农村领域为例 [J]. 情报工程 , 2026, 12(2): 117-126.

doi:10.3772/j.issn.2095-915x.2026.02.010

3. Library of Capital Normal University, Beijing 100048, China;
4. China Agricutural University Library, Beijing 100083, China

Abstract: [Objective/Significance] High-quality peer review is an important prerequisite for selecting high-level science and technology projects, the quality of peer review at the project initiation stage has an important impact on the allocation of scientific research funding resources. [Methods/Process] Based on the evaluation data of 184 projects from 8 types of national key R&D projects in the field of agriculture and rural areas undertaken by China Rural Technology Development Center from August to October 2023, this study compared and analyzed the peer review reliability of proposed projects and unapproved projects. [Results/Conclusions] It is found that the reliability coefficients of individual experts and the whole group of experts of all projects are high, and the scoring consistency of the proposed project is generally higher than that of the unapproved project, and the reliability coefficients of different fields are different. Based on this, this study puts forward policy suggestions such as further clarifying the review criteria, optimizing the selection of experts, focusing on areas with large outliers, analyzing the reasons for the outliers of uninitiated projects, and continuously improving the review system, so as to promote the scientific and standardized development of peer review of science and technology projects.
Keywords: National Key Research and Development Plan; Agriculture and Rural Areas; Peer Review; Quality

引言

国家重点研发计划是以解决国家重大需求或关键科学问题为目标导向，根据任务需求由目标鲜明、界限清晰的项目群作为实施载体开展研究的科技计划[1]。作为国家科技计划管理改革的重要成果，其整合了原有的973计划、863计划、国家科技支撑计划、国际科技合作与交流专项，以及相关部门管理的公益性行业科研专项[2]。该计划旨在破解制约我国科技创新发展的深层次问题，提高财政资金使用效益，遵循科技发展的客观规律，激发科技人员的积极性和创造力，进一步推动高水平的科技自立自强。自2016年启动以来，国家重点研发计划受到了科技界的广泛关注。

高质量的同行评议是确保遴选出高水平科技项目的重要前提。立项阶段的同行评议质量直接影响科研基金的配置方向及研究质量，并最终决定科技项目资助效果[3]。学术界围绕科技项目同行评议信度展开了大量研究，主要集中在两个方面：一是评估同行评议专家的评审信用，如王国彪等[4]探讨了国家自然科学基金同行评议结果的评价方法，并分析了专家遴选的因素，提出了项目熟悉度、项目资助度和项目综合评价指标偏差三个定量评价指标。陈媛等[5]从科研活跃程度、评审业绩和工作态度等方面建立了专家信用评价体系。贺晓宇等[6]提出了一个包含资质能力、行为规范、评审表现的评审专家信用评价指标体系。刘润达等[7]则聚焦于科研项目评审活动中专家诚信建设的问题，通过对比欧美与中国的实践，提出了保障评审专家诚信的多项建议。二是对科技项目同行评议专家打分的可靠性进行数理分析，如马宗文等[8]选取“战略性科技创新合作”重点专项两批次项目的评审数据，深入研究了146名同行评议专家的691份评审结果。Jerrim等[9]

分析了 4000 份社会科学资助申请和 15000 份评审报告的数据，研究了同行评审的一致性、提名评审的影响以及单一负面评审对资助结果的影响。Gallo 等 [10] 梳理了美国生物科学研究所 2063 份申请书的同行评议数据，探究了专家打分质量与后期项目产出的潜在关系。此外，还有研究深入探讨了美国国立卫生研究院科技项目 [11]、欧盟玛丽居里科技项目 [12] 的评审可靠性。

综合来看，现有研究探讨了科技项目同行评议的信度，为本研究提供了重要的理论借鉴和方法参考。然而，大多数研究聚焦于已立项的科技项目，较少关注未立项项目，鲜见对国家重点研发计划已立项与未立项同行评议信度进行对比分析。因此，本文基于农业农村领域 8 个国家重点研发计划同行评议打分数据，以同行评议信度为研究对象，进行拟立项与未立项项目的评审信度对比分析，希望为国家重点研发计划的监督管理提供决策参考，为科技项目同行评议相关领域的研究提供方法借鉴。

1 研究设计

1.1 研究思路

本文聚焦农业农村领域国家重点研发计划重点专项同行评议结果的信度分析，在借鉴已有研究基础上，采取以下研究思路：首先，统一计算同行评议的信度系数，评估国家重点研发计划重点专项项目专家打分的可靠性。其次，分别计算拟立项与未立项项目专家打分的离散度与异常值，并进行对比分析。最后，综合比较信度系数、离散度和异常值分布，深入分析拟立项与未立项项目专家打分之间的差异，并探究可能的原因，基于此，提出针对性的政策建议，希望为优化国家重点研发计划重点专项项目同行评议机制提供参考借鉴。

1.2 数据与方法

2023 年 8—10 月，中国农村技术开发中心顺利完成农业农村领域 8 个国家重点研发计划重点专项项目评审工作。本研究数据来源于本次评审工作中记录的专家打分情况，共包含 120 名专家对 184 个项目的 2651 份评审结果，其中拟立项项目 76 个，未立项项目 108 个，具体如表 1 所示。每一组项目评审中分别有 15 名专家，在使用专家打分时，并未去掉最高分和最低分，而是采用了全部专家打分数值。在已有研究 [8, 13] 的基础上，本文设计了同行评议打分质量的综合计算方法，包括信度系数、离散度、异常值三项指标。

表 1 项目评审基本情况

项目序号	拟立项数量	未立项数量	评审专家数量
第 1 组	9	17	15
第 2 组	5	8	15
第 3 组	15	19	15
第 4 组	10	18	15
第 5 组	12	16	15
第 6 组	12	13	15
第 7 组	10	12	15
第 8 组	3	5	15

1.2.1 信度系数计算方法

信度系数是衡量测量工具（如问卷、评分等）可靠性的指标，它反映了测量结果的一致性和

doi:10.3772/j.issn.2095-915x.2026.02.010

稳定性[14]。单个专家评分信度系数计算公式为：

$$r=\frac{\Sigma_i\Sigma_k l_{jk}}{\Sigma_j\Sigma_k\frac{(l_{jj}+l_{kk})}{2}}$$

其中，l_{jj} 表示第 j 个专家对所有项目评分的离均差平方和，即第 j 个专家的评分与所有项目平均得分之差的平方的总和。l_{kk} 表示第 k 个专家对所有项目评分的离均差平方和，同理于 l_{jj}。l_{jk} 表示第 j 个专家对某一项目 i 的评分与该项目平均得分之间的差和第 k 个专家对同一项目评分与该项目平均评分之间的差的乘积。即两个专家对同一项目评分差异的乘积，反映了他们评分的一致性。

整组专家评分信度系数使用斯皮尔曼 - 布朗公式：

$$R=\frac{m\bar{r}}{1+(m-1)\bar{r}}$$

其中，m 代表评分的专家个数，$\bar{r}$ 表示单个专家评分的信度系数均值。

1.2.2 离散度计算方法

离散度（Divergence）是向量微积分中的一个概念，用于描述向量场在某一点处的发散程度[15]。离散度的计算方法如下：首先计算每个分量的偏导数，分别计算向量场 F 在 x、y 和 z 方向上的分量 Fx、Fy 和 Fz 关于 x、y 和 z 的偏导数。其次求和，将这些偏导数相加，得到离散度。本文中，同一项目专家打分的离散度数值通过 EXCEL 中的函数 AVEDEV 计算得出。

1.2.3 异常值计算方法

在异常值检验中，本文使用 Grubbs 检验和 Dixon 检验两种方法，它们是两种常用的异常值检测方法，在统计学中被广泛应用[16]。

Grubbs 检验，也称为 ESD（Extreme Studentized Deviate）检验，是一种用于识别单变量数据集中单个异常值的统计方法，要求数据集服从正态分布。Grubbs 临界值取决于显著性水平（α）和样本大小（n），可以通过查表获得。如果发现异常值，将其排除后，对剩余数据重复上述步骤，直到没有更多的异常值被发现。Grubbs 检验的步骤如下：①将数据集从小到大排序。②计算数据集的均值（mean）和标准差（standard deviation）。③确定最大值或最小值与均值的差距，较大的那个值被视为可疑值。④计算可疑值的 z-score（标准分数），如果 z-score 大于 Grubbs 临界值，则该值被视为异常值。

Dixon 检验，也称为 Q 检验，是一种用于检测数据集中单个异常值的方法，特别适用于小样本数据集。Dixon 检验的临界值取决于显著性水平（α）和样本大小（n）。与 Grubbs 检验不同，Dixon 检验不要求数据服从正态分布。Dixon 检验的步骤如下：①将数据集从小到大排序。②确定最大值（或最小值）与前一个值（或后一个值）的差值，以及最小值（或最大值）与后一个值（或前一个值）的差值。③计算 Q 统计量，其定义为最大值与前一个值的差值与最小值与后一个值的差值的比值。④将计算出的 Q 值与 Dixon 表中的临界值进行比较，如果 Q 值大于临界值，则最大值（或最小值）被视为异常值。

2 分析结果

2.1 专家评分信度分析

表 2 展示了 8 个国家重点研发计划重点

专项同行评议专家打分的单个专家信度系数和整组专家信度系数。其中单个专家信度系数，反映了某一位专家对某个项目评价的一致性，数值越接近 1，表示该专家的评价越一致，信度越高。整组专家信度系数，反映了整组专家对某个项目评价的一致性，同样数值越接近 1，表示整组专家的评价越一致，信度越高。

首先，在单个专家信度系数方面，所有领域的单个专家信度系数都超过了 0.7，这通常被认为是可接受的信度水平。其中，第 7 组的单个专家信度系数最高，达到了 0.947，这表明该领域单个专家的评价具有较高一致性。而第 6 组的单个专家信度系数相对较低，为 0.772，表明在该领域单个专家的评价存在一定的差异性。

其次，在整组专家信度系数方面，所有领域的整组专家信度系数都非常高，普遍在 0.98 以上，这表明专家组内的评价具有很高的一致性。其中，第 7 组的整组专家信度系数最高，达到了 0.996，这显示了该领域专家评审的极高一致性。而第 6 组的整组专家信度系数相对较低，为 0.981，尽管如此，这个数值仍然表明了专家组评价的高一致性。

综合来看，单个专家信度系数的范围从 0.772~0.947，而整组专家信度系数则在 0.981~0.996 之间。无论是单个专家还是整组专家，所有项目的信度系数都相对较高，这表明专家评审在这些领域内具有较高的一致性和可靠性。此外，不同领域的信度系数存在差异，这可能与不同领域的特性、专家的专业知识水平以及评价标准的明确性有关。

表 2　同行评议专家打分信度系数

项目序号	单个专家信度系数	整组专家信度系数
第 1 组	0.811	0.985
第 2 组	0.811	0.985
第 3 组	0.906	0.993
第 4 组	0.856	0.989
第 5 组	0.872	0.990
第 6 组	0.772	0.981
第 7 组	0.947	0.996
第 8 组	0.880	0.991
平均值	0.857	0.989

2.2　专家评分离散度分析

表 3 展示了 8 个国家重点研发计划重点专项同行评议专家打分最大值、最小值、平均值和离散度数值。首先，从最大值和最小值来看，在拟立项项目中，第 3 组的最大值最高，为 99，而最小值最低的是第 2 组，为 70。这表明在评议中，第 3 组得到了较高的评价，而第 2 组则存在较大的评价差异。在未立项项目中，第 6 组的最大值最高，为 99，但最小值也相对较低，为 58。比较而言，虽然拟立项与未立项项目的最大值评分相差不多，但在最小值方面，拟立项项目的评分普遍高于未立项项目。

其次，从平均值来看，拟立项项目的平均打分普遍高于未立项项目，这表明评议专家对拟立项项目的整体评价更为积极。具体来看，第 3 组和第 8 组的平均打分最高，分别为 87.88 和 89.54，这可能反映了这些领域的项目在同行评议中得到了较高的认可。相反，第 6 组和第 5 组的平均打分相对较低，分别为 85.78 和 86.89，这可能意味着这些领域的项目在评议中存在一定的争议。

doi:10.3772/j.issn.2095-915x.2026.02.010

最后，从离散度来看，离散度较低意味着专家打分较为一致，而离散度较高则表明打分差异较大。在平均离散度方面，拟立项项目的平均离散度为 4.26，而未立项项目的平均离散度为 5.03，总体上拟立项项目的评分一致性高于未立项项目。在具体离散度方面，拟立项项目中离散度最低的是第 8 组（2.91），最高的是第 2 组（4.87）；未立项项目中离散度最低的是第 3 组（4.46），最高的是第 8 组（5.51）。拟立项项目离散度最低和未立项项目离散度最高均为第 8 组，二者之间形成鲜明对比。

表 3　同行评议专家打分离散度数值

项目类型	拟立项项目				未立项项目			
	最大值	最小值	平均值	离散度	最大值	最小值	平均值	离散度
第 1 组	93	71	84.14	4.03	94	55	74.98	4.51
第 2 组	97	70	86.48	4.87	96	55	76.34	4.83
第 3 组	99	72	87.88	4.47	98	55	74.21	4.46
第 4 组	98	70	88.32	4.79	97	55	76.29	5.37
第 5 组	97	73	86.89	4.38	96	45	75.6	5.04
第 6 组	99	72	85.78	5.08	99	58	77.03	5.92
第 7 组	98	75	85.34	3.53	91	60	73.36	4.63
第 8 组	97	73	89.54	2.91	92	53	75.54	5.51
平均值	\	\	86.80	4.26	\	\	75.42	5.03

2.3　专家评分异常值分析

表 4 展示了 8 个国家重点研发计划重点专项同行评议专家打分异常值检验结果，Grubbs 方法检验出异常值 37 个，其中拟立项项目 9 个，未立项项目 28 个；Dixon 方法检验出异常值 56 个，其中拟立项项目 15 个，未立项项目 41 个。

首先，从拟立项与未立项的异常值数量上来看，未立项项目在两种检验方法下识别出的异常值数量均高于拟立项项目，分别为 28（Grubbs）和 41（Dixon），而拟立项项目则分别为 9（Grubbs）和 15（Dixon），表明未立项项目在评审过程中存在更多的评分不一致性，或者这些项目本身在质量上存在较大差异，导致评审专家的打分出现较大波动。具体到各个领域，可以看到在拟立项项目中，第 4 组和第 8 组的异常值数量相对较多，在未立项项目中，第 3 组和第 4 组的异常值数量相对较多。表明在这些领域的项目评审中，专家的意见可能存在一定分歧。

其次，从 Grubbs 检验与 Dixon 检验对比分析角度来看，Dixon 检验在大多数领域识别出的异常值数量都高于 Grubbs 检验，表明 Dixon 检验对于非正态分布的数据更为敏感，能够检测出更多的异常值。在拟立项项目中，Dixon 检验在第 2 组、第 4 组和第 8 组识别出的异常值数量明显高于 Grubbs 检验，在未立项项目中，两种检验方法在第 3 组和第 4 组均识别出较高的异常值数量，这表明在这些领域专家打分可能偏离正态分布。

总的来看，未立项项目相比拟立项项目存

在更多的异常值，Dixon 检验相比 Grubbs 检验检测到更多的异常值，在具体领域中，第 4 组在评议过程中存在较多的极端评分，在后期的项目评审管理过程中需要加以重点关注。

表 4　同行评议专家打分异常值检验

项目类型	拟立项项目		未立项项目	
	Grubbs 检验异常值数量	Dixon 检验异常值数量	Grubbs 检验异常值数量	Dixon 检验异常值数量
第 1 组	0	0	5	8
第 2 组	2	4	3	1
第 3 组	0	1	5	8
第 4 组	4	7	6	8
第 5 组	0	2	4	7
第 6 组	0	0	1	4
第 7 组	1	0	3	4
第 8 组	2	1	1	1
合计	9	15	28	41

2.4　拟立项与未立项结果对比分析

表 5 展示了 8 个国家重点研发计划重点专项中拟立项与未立项同行评议打分的对比情况。首先，离散度对比分析显示，拟立项项目的平均离散度普遍低于未立项项目，即拟立项项目的专家打分更为一致。具体到各个领域，第 3 组、第 8 组在拟立项项目中平均打分最高，而第 6 组和第 5 组相对较低。未立项项目中，第 6 组的最大值与拟立项项目相差无几，但最小值显著低于拟立项项目，这表明未立项项目可能在质量上存在较大差异。

其次，从异常值对比分析来看，异常值揭示了拟立项与未立项项目在评审过程中的评分波动情况。未立项项目在 Grubbs 和 Dixon 两种检验方法下识别出的异常值数量均高于拟立项项目，表明未立项项目在评审过程中存在更多的评分不一致性。具体到各个领域，第 4 组在评议过程中存在较多的极端评分，无论是拟立项还是未立项项目，意味着该领域项目的专家评分存在较大差异。

表 5　拟立项与未立项同行评议专家打分对比分析

项目类型	拟立项项目		未立项项目	
	离散度	Dixon 检验异常值数量	离散度	Dixon 检验异常值数量
第 1 组	4.03	0	4.51	8
第 2 组	4.87	4	4.83	1
第 3 组	4.47	1	4.46	8
第 4 组	4.79	7	5.37	8
第 5 组	4.38	2	5.04	7
第 6 组	5.08	0	5.92	4
第 7 组	3.53	0	4.63	4
第 8 组	2.91	1	5.51	1

最后，结合信度系数来看，拟立项与未立项项目在单个专家和整组专家的信度系数上均表现出较高的一致性。这表明无论是单个专家还是整组专家，对项目的评审都具有较高的可靠性。然而，具体到不同领域，第 7 组的单个专家信度系数最高，而第 6 组相对较低。整组专家信度系数方面，第 7 组同样最高，第 6 组次之。

doi:10.3772/j.issn.2095-915x.2026.02.010

3 结论与建议

3.1 研究结论

本文首先分析了 8 个国家重点研发计划重点专项的单个专家信度系数和整组专家信度系数，其次分析了拟立项与未立项项目的离散度和异常值，最后进行了对比分析，主要研究结论如下：

（1）农业农村领域国家重点研发计划项目同行评审信度系数较高。所有项目的单个专家信度系数和整组专家信度系数均较高，8 个国家重点研发计划重点专项的整组专家信度系数更是高达 0.98 以上，充分证明专家评审具有较高的一致性和稳定性。

（2）拟立项项目的评审可靠性普遍高于未立项项目。拟立项项目评分一致性更高，平均离散度低于未立项项目，专家打分更为一致。同时，与拟立项项目相比，未立项项目异常值更多，在 Grubbs 和 Dixon 两种检验方法下识别出的异常值数量均高于拟立项项目，在评审过程中存在更多的评分不一致性，在后期的项目评审管理过程中需要给予更多关注。

（3）不同项目间信度系数存在明显差异。不同领域的信度系数存在显著差异，如第 7 组单个专家信度系数高达 0.95，而第 8 组单个专家信度系数为 0.88。这可能与领域特性、专家专业知识水平以及评价标准的明确性有关。

3.2 政策建议

（1）进一步明确国家重点研发计划项目同行评议评审标准。评审标准是同行评议的核心依据，其明确性直接关系到评审结果的公正性和一致性 [17]。本文发现，在部分项目评审打分中，存在信度系数相对较低、评分离散度较大、异常值数量较多的情况。针对这些项目，应深入剖析该项目的特性和评审难点，组织领域内的专家学者和项目管理人员，共同研讨并细化评审标准 [18]。明确的评审标准能够为专家提供清晰的评价指引，使不同专家在面对同一项目时，能够基于统一的标准进行客观、公正的评价，从而提高评审结果的可靠性和可信度。

（2）重点关注国家重点研发计划项目异常值较多的领域。异常值的出现往往意味着评审过程中存在某些不确定因素或项目本身存在问题，对于异常值较多的领域应加强评审过程的监督和管理 [19]。本文建议建立异常评分复核机制，对于 Grubbs 和 Dixon 检验中识别的异常值进行多级复核，例如由领域首席专家或独立第三方委员会对异常评分项目进行二次评审，确保评分偏差源于项目质量而非评审疏漏。同时，可以开发智能化异常检测工具，结合大数据分析技术，构建动态评分监控系统，实时识别极端评分并触发预警，辅助评审管理人员快速响应。此外，还可以考虑引入国际科技项目评审经验，如欧盟“玛丽·居里计划”的评审流程，在争议性项目中增设专家讨论环节，通过辩论达成共识后再进行终审 [12]。

（3）分析国家重点研发计划未立项项目异常值原因。未立项项目的评分不一致性可能隐藏着影响评审质量的关键因素。针对未立项项目存在的评分不一致性，应深入分析其原因。首先，可以通过对评审数据的挖掘和分析，找

出评分差异较大的项目和专家，了解专家的评分依据和理由，是否存在对项目理解的偏差或评价标准的不一致。其次，可以组织专家研讨会，针对未立项项目的评审情况进行交流和讨论，分享评审经验和观点，促进专家之间的共识形成。此外，还可以考虑增加专家讨论环节，在项目评审过程中，让专家就项目的创新性、可行性、预期效益等关键问题进行充分讨论，通过思想碰撞提高专家对项目的整体认知，减少因个体差异导致的评分不一致，从而提高评审的准确性。

（4）持续优化国家重点研发计划项目评审体系。评审体系的有效性是确保国家重点研发计划项目同行评议质量的关键。本研究建议国家重点研发计划项目同行评议管理机构建立评审体系的动态评估机制，定期跟踪和评估评审体系的有效性。一方面，要不断优化评审流程，简化不必要的环节，提高评审效率。例如，可以探索建立线上评审系统，实现项目申报、专家评审、结果反馈等全流程的信息化管理，减少人工操作的误差和时间成本。另一方面，要持续改进评审方法，引入新的评审技术和工具，如人工智能辅助评审、大数据分析等，为评审工作提供更科学、更精准的决策支持[20]。同时，要定期收集和分析评审数据，包括专家评分数据、项目立项情况、项目执行效果等，通过数据分析及时发现评审过程中存在的问题，如专家信度系数的变化趋势、异常值的分布规律等，并据此采取相应的改进措施，确保评审工作的持续改进，不断提升整个评审体系的可靠性和公正性，为国家重点研发计划项目的高质量发展提供坚实的保障。

参 考 文 献

[1] 宋欢，陈金辉 . 国家重点研发计划专项专家库建设存在的问题与对策 —— 基于一线项目管理者的访谈研究 [J]. 中国科技论坛，2024(3): 50-59.

[2] 杨毅，魏瑞芝，张依，等 . 国家重点研发计划资助项目空间分布研究与启示 [J]. 科技进步与对策，2019, 36(14): 25-30.

[3] 张璐杰 . 国家自然科学基金项目立项同行评议质量控制研究 [D]. 北京 : 北京科技大学，2017.

[4] 王国彪，彭芳瑜 . 国家自然科学基金同行评议结果评价方法与专家遴选因素分析 [J]. 中国科学基金，2008(6): 372-376.

[5] 陈媛，樊治平，谢美萍 . 科研项目同行评议专家水平的评价研究 [J]. 科学学与科学技术管理，2009, 30(10): 38-42.

[6] 贺晓宇 . 基于 TOPSIS 的科技项目评审专家信用评价模型研究 [J]. 科技管理研究，2020, 40(3): 32-38.

[7] 刘润达，赵星宇，钟伟，等 . 科研项目评审活动中专家诚信建设刍议 [J]. 中国科学基金，2024, 38(5): 885-890.

[8] 马宗文，张金倩楠 . 国家重点研发计划项目立项同行评议可靠性分析 [J]. 科学学研究，2025, 43(4): 1-15.

[9] JERRIM J, DE VRIES R. Are peer reviews of grant proposals reliable? An analysis of Economic and Social Research Council (ESRC) funding applications[J]. Social Science Journal, 2023, 60(1): 91-109.

[10] GALLO S A, CARPENTER A S, IRWIN D, et al. The validation of peer review through research impact measures and the implications for funding strategies[J]. Plos One, 2014, 9(9): e106437.

[11] EROSHEVA E A, GRANT S, CHEN M C, et al. NIH peer review: criterion scores completely account for racial disparities in overall impact scores[J]. Science Advances, 2020, 6(23): eaba4868.

[12] PINA D G, BULJAN I, HREN D, et al. A retrospective analysis of the peer review of more than 75,000 Marie Curie proposals between 2007 and

doi:10.3772/j.issn.2095-915x.2026.02.010

2018[J]. elife, 2021, 10: e59338.

[13] 黄正南 . 专家评判的信度分析 [J]. 中国卫生统计 , 2000(3): 27-29.

[14] 田雪垠 , 郑蝉金 , 郭少阳 , 等 . 基于多层验证性因素分析的各种信度系数方法 [J]. 心理学探新 , 2019, 39(5): 461-467.

[15] 王宝金 , 李学良 . 科技期刊评价指标的离散度分析 [J]. 中国科技期刊研究 , 2007, 18(1): 81-83.

[16] 刘伟力 , 姜箴 , 丁春光 , 等 . 两种统计检验法在异常评分筛选中的应用研究 [J]. 电子产品可靠性与环境试验 , 2020, 38(S1): 16-19.

[17] 李曼迪 , 赵筱媛 , 刘晓娟 . 颠覆性技术项目中同行评议机制改进思路与国际实践 [J]. 中国科技论坛 , 2021(12): 182-188.

[18] 戴贵宝 , 张斌 , 姚娃丽 , 等 . 科技计划项目视域下的同行评议研究：文献综述和研究展望 [J]. 全球科技经济瞭望 , 2024, 39(8): 17-24.

[19] 王艳芳 , 白波 . 国外科技项目评审调研及对我国的启示 [J]. 技术与创新管理 , 2012, 33(3): 289-292.

[20] 李东 , 郝艳妮 , 彭升辉 , 等 . 人工智能技术在科学基金项目管理中的实践与探索 [J]. 中国科学基金 , 2022, 36(5): 790-797.

（责任编辑：浦墨）