

对同行评议专家的反评估分析

赵黎明

徐孝涵 张卫东

(天津大学管理学院) (国家自然科学基金委员会)

国家自然科学基金委员会成立八年来,根据国家发展科学技术的方针、政策、规划,有效地运用科学基金,指导、协调和资助基础研究和部分应用基础研究工作,在发现和培养人才,促进科学技术进步和经济、社会发展方面做出了重要的贡献。

基金资助项目的评审采用同行评议方法,该方法是基础研究领域内进行项目评审、检查和评估成果,优秀人才及学术团体的一种有效而可行的方法,被广泛用于科研管理,高等人才的培养,成为科技投资决策的辅助手段,促进了科研管理的科学化、民主化进程。

但由于基础研究领域十分复杂,同行评议方法本身又具有一定局限性,近年来,受到国内外学术界一些专家的尖锐批评。国内外学术界对同行评议方法的批评,从实质上看,很多是针对同行评议专家的。因此对同行评议专家进行反评估,提高同行评议专家队伍水平是当前亟待解决的重要课题。

国家自然科学基金委员会聘请大量同行评议专家对申请项目进行评审,以期获取有关申请项目的最大量的信息,使最终就项目是否给予资助作出准确的决策,这就要求专家们对此项工作具有严肃认真、公正负责的态度和较高的学术水平、准确的判断能力。同行评议专家以往进行的评议工作情况,在一定程度上反映了同行评议专家的水平 and 态度,将同行评议专家以往评议工作状况用有关指标进行分析,对其进行反评估,这对于同行评议专家的挑选、调整是十分必要的,而且可使同行评议专家库处于动态优化过程中。

1 评议项目累计数

评议项目累计数是指某评议专家曾评审过的国家自然科学基金项目数的总和。我们认为累计数可以较好地反映评议专家评议工作经验的多少。而经验的多少,会对其评议结果产生较大的影响。另外,累计数的大小在一定程度上反映出某评议专家以往工作业绩的好坏。当某评议专家以往的评议工作认真负责,得到基金管理人员的重视,而被经常予以聘用,其评议项目就增多、累计数加大。因此累计数可以间接反映基金管理人员对评议专家工作的评价,从一个侧面反映了评议效果的优劣。在同一专业领域内,累计数指标的作用更为明显。

2 评议结果离散率

评议结果离散率分为:横向离散率,表明某评议专家与其他评议专家对项目评价时产生的认识上的差异大小;纵向离散率,反映的是某评议专家在历次评议中对所评的项目在认识上的波动性的大小。

通过收集统计各评议专家的历次评议结果,并进行初步整理后可得下列矩阵:

甲
乙
甲
乙

中国科学基金 95.1

$$[x] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}$$

式中: x_{ij} —评审专家 j 对第 i 个项目的评议值。

在原始数据阵 $[X]$ 中, 令

$$a_{ij} = \frac{1}{m} \sum_{k=1}^m (X_{ik} - X_{jk})^2 \quad \begin{cases} i = 1, 2, \cdots, n \\ j = 1, 2, \cdots, m \end{cases}$$

分别求出每个专家关于每个项目的横向离散率, 可以得到横向离散率矩阵。当横向离散率矩阵中元素值 a_{ij} 均较大时, 说明整个评议专家组在对项目的认识上存在较大的分歧。此时基金管理人员应对差异性大的项目进行专门研究, 看是否是由于革命性创新造成的非共识问题所致, 如果是此类问题、应另案解决。排除了这种可能之后, 则应从专家的学术水平、认识能力、道德修养 (如公平性) 等方面予以考虑。

在原始数据阵 $[X]$ 中, 令

$$b_{ij} = \frac{1}{n} \sum_{k=1}^n (X_{ik} - X_{kj})^2$$

式中: b_{ij} ——评议专家 j 关于项目 i 的纵向离散率。

分别求出每个专家关于每个项目的纵向离散率, 可以得到纵向离散率矩阵。由纵向离散率矩阵可以描绘出各专家评议结果纵向离散率分布情况, 根据我们对部分学科现有资料的数据分析, 所得曲线不尽相同, 但大致归纳为三类。

当曲线处于波浪状态时, 一般来说是正常的。学科管理人员在分派评议项目时, 一般只考虑评议专家与拟评项目的研究方向是否一致和评议专家能够承担的评议工作量, 一般不考虑项目本身的优劣。因此同一专家对不同项目有不同的评议结论, 是不难理解的。从统计规律而言, 评议结果将呈一定规律性的波动。当然专家对资助比例的理解也是造成这一波动的原因。

当曲线为直线时, 一般来说是不正常的。从曲线来看某同行评议专家对其所评项目的评议值没有区别。但是不同项目的立论依据, 研究方案、研究基础, 研究队伍的不同造成了申请项目事实上的优劣之别。同时专家本人对项目的认识程度也有差别。造成这种情况出现的原因可能是多方面的: 由于该评议专家对评议工作缺乏严肃认真态度, 不负责任, 采取一律通过或一律否决的作法, 随意打分造成的; 也可能是专家评审时存在正负宽大化误差, 即所得结果比被考核者实际应得结果或高或低的倾向; 也可能是专家判断时存在中心化倾向的误差, 即由于评议专家认为自己对被评议者的项目不甚了解, 避免使用“极好”、“极坏”的极端词句, 使评议结果集中于中间状态。不论何种原因都不符合遴选项目的评审要求。

当本不正常的直线偶然发生大幅度跳跃时, 反映该专家对大多数项目的评分接近, 但只有个别项目的评分出现大幅度跳跃。这种情况应引起注意。可能其中存在不公正的人为因素; 另外, 也可能是某个重大革命创新性项目被评议专家发现。

横向及纵向离散率分析从不同角度在一定程度上反映了评议专家的评议工作状况, 但是

在综合评价各专家水平与公正性时,要消除不同专家因掌握评议标准不同而带来的影响,需要进行综合离散率分析。

综合离散率的计算是原始数据阵 $[X]$ 的基础上,令

$$X'_{ij} = \left(X_{ij} - \frac{1}{m} \sum_{k=1}^m X_{ik} \right)$$

式中: X'_{ij} ——评议专家 j 与整个评议小组在项目 i 认识上的差异性。

由此得到偏差距阵 $[X']$ 。 $[X']$ 中第 j 列的平均值为:

$$\bar{x}'_j = \frac{1}{n} \sum_{i=1}^n x'_{ij}$$

$\bar{x}'_j$ ——第 j 专家历次评议结果与整体评议值偏差的平均值。令

$$C_j = \frac{1}{n} \sum_{i=1}^n (x'_{ij} - \bar{x}'_j)^2$$

可得下列一维矩阵

$$[C] = [c_1, c_2, \dots, c_j, \dots, c_m]$$

式中: c_j ——第 j 个评议专家的综合离散率。

在推导 c_j 时,因以专家 j 对项目 i 的评议值与所有专家对该项目的平均值的偏差 x'_{ij} 代替 x_{ij} 进行分析,而 x'_{ij} 值的大小不受各专家对评议项目的标准掌握不同的影响,因此最后所得到的 c_j 已消除了评议专家因掌握标准不同而带来的差异性,将所得结果用于对专家评议工作情况评估,就更具有较为实用的价值。

对于 $[C]$ 中的较大元素,应该进行分析。例如:当 $[C]$ 中最大元素为 C_k 时、应该对专家 K 予以注意。首先应该寻找是哪些项目的评审值差距过大而造成了 C_k 偏大。寻找方法如下:令 $D_{ij} = (x'_{ij} - \bar{x}'_j)^2$, 得到矩阵 $[D]$ 。可根据 D_{ij} 的大小找到有疑问的项目进行分析,要判断其是否由于项目具有革命性创新而造成“非共识”问题,对于这类问题,应组织有关专家进一步评审,在仍未达成一致意见时,可拨出预研基金予以资助,以保证具有革命性创新的项目得到支持,保证专家的真知灼见不被埋没。在排除了这类问题之后,分析 C_k 偏大的原因,大致为主客观两个方面:从客观因素来看,由于该专家本身水平有限,不能正确认识申请项目的本质,从而不能对其进行正确评价;也可能是由于评审项目与专家专业不尽相同、专家勉为其难所致。从主观因素来看,该专家具有较高评价水平,但由于主观上对申请项目有偏见,而不能对其进行客观合理的评价,从一个方面反映出专家的不公正性。

3 评议结果命中率

命中率是指在同行评议中,某评议专家对申请项目的评议结果与项目是否获资助的一致性程度的大小。

命中率

$$P_j = (x_{0j}/x_{1j}) \times 100\%$$

x_{0j} ——评议专家的 j 的评议结果与最终评议结果相一致的项目数;

x_{1j} ——评议专家 j 评议项目数;

分别求出每一专家的命中率,可得到下列评议专家命中率矩阵:

$$[P] = [P_1, P_2, \dots, P_j, \dots, P_m]$$

使用该矩阵可进行各专家间的相对比较。

基金委成立以来资助了大量科研项目,根据调查结果表明,绝大部分受资助项目取得了不同程度的成果,对我国科技事业发展起到了重要的作用。可以说,资助是有效的。从这个角度来看同行评议专家的评议结果是否与基金委资助的决策相符,是对同行评议专家评议能力的一种检验。同时评议专家政策水平高,深刻了解基金资助宗旨;学术水平高,准确判断申请项目的价值;同时认真负责,公正合理,那么所评议项目的命中率一定高。

当然,命中率的高低还与该专家一次评议项目的多少有关。通常评议项目数多的专家的命中率要高于评议项目数少的专家。一次评议项目较多,可以使该专家在较大范围内进行比较,从而找出其中最优的方案。但从工作量角度看,建议每个专家评审的项目数不超过10项。

4 资助项目成功率

在基金数额有限的情况下,我们希望每项资助都能获得成功,促进我国科技事业的发展,这也是检验基金工作的重要标准。但是,有些资助项目可能会由于这样或那样的原因而最终没有获得成功。这从一个侧面反映了同行评议结果的正确与否,因此有必要提出成功率的概念以作为评议专家反评估分析的重要基准之一。

成功率(r_j)是指实践证明取得成功的资助项目与总资助项目的比率

$$r_j = (r_{0j}/r_{1j}) \times 100\% \quad (j = 1, 2, \dots, m)$$

r_{0j} —评议专家 j 认为可以获得资助,且经实践证明是成功的项目的个数;

r_{1j} —评议专家 j 认为应获得资助的项目数。

分别求出每一专家评议工作的成功率,可得下列成功率矩阵:

$$[r] = [r_1, r_2, \dots, r_j, \dots, r_m]$$

使用该矩阵可进行各专家间的相对比较。

资助项目成功率是从最终结果对同行评议专家的评议水平进行总体反映的一个指标。尽管一个项目能否成功,受到诸多内外因素的影响,但是专家根据自己所掌握的国内外研究状况,申请表提供的有关信息,运用自己的研究和判断能力,是可以对项目的成功概率做出相对准确的判断。因此成功率的高低,从本质上反映了同行评议专家评议水平的高低。较高的成功率反映了同行评议系统能有效地实现基金资助的宗旨。因此它们与其他指标相比,具有更强的说服力,是评价指标体系中非常重要的一个指标。但是,作为基础研究项目,其成功与否不是短时间能够反映的,而且成功的标准是一个涉及面极广,非常复杂的问题,我国对这类问题的研究尚需深入,这就为成功率的判断带来较大困难。随着管理工作不断科学化,这一指标的计算将成为现实。

上面我们从累计数、离散率、命中率、成功率等不同角度对同行评议专家的评议情况进行了分析,需要指出的是:

(1) 这些指标之间是有联系的,对指标分开单独进行评价是不全面的,因而有必要对专家的评议工作业绩进行综合分析,并根据每个指标的重要性而给出权重。

(2) 由于同行评议专家评审系统运行复杂,项目千差万别,考虑到可操作性,我们所列指标是对评议结果的统计分析,不可能也没有必要对数万专家进行逐个定性分析。因此所得结果应该而且也只能作为基金管理人员对同行评议专家进行反评估时参考,不应作为唯一不可改变的结论。

最后, 应该指出尽管学术界对同行评议在项目评审, 尤其是非共识项目评审中的作用有所争议, 但是根据我们掌握的大量调查资料 and 我国实行科学基金制 12 年走过的路程来看, 除为数不多的非共识项目 (革命性创新项目) 之外, 对于基金申请中的绝大多数共识和非共识项目, 考虑多数评议专家意见的作法是可行的。因此, 我们可以根据评议结果的统计值为反评估提供依据。

APPRAISAL AND ANALYSIS OF EXPERTS' PEER REVIEW

Zhao Liming

(School of Management, Tianjin University, Tianjin 300072)

Xu Xiaohan Zhang Weidong

(National Natural Science Foundation of China, Beijing 100083)

上接第 75 页

序 号	姓 名	专 业	所 在 单 位
23	张学光	肿瘤学	苏州医学院
24	夏照帆 (女)	临床医学基础	第二军医大学
25	左萍萍 (女)	神经药理学	中国医学科学院基础医学研究所
26	姚檀栋	冰川气候与环境	中国科学院兰州冰川冻土研究所
27	丁仲礼	第四纪地质学	中国科学院地质研究所
28	史晓颖	地层学及古生物学	中国地质大学 (北京)
29	陈 骏	地球化学	南京大学
30	朱日祥	地磁学	中国科学院地球物理研究所
31	李国营	地球物理	中国科学院测量与地球物理研究所
32	翟世奎	海洋地质学	中国科学院海洋研究所
33	张志东	金属材料	中国科学院金属研究所
34	沈保根	磁学	中国科学院物理研究所
35	李晓光	固体物理	中国科学技术大学
36	倪晋仁	泥沙学	北京大学
37	陆祖宏	生物电子学	东南大学
38	张晨曦	计算机	国防科技大学
39	陈晓波	凝聚态物理	南开大学
40	王鹏业	光学	中国科学院物理研究所