

·基金纵横·

# 国家自然科学基金专家评议状况评估初探

谷瑞升<sup>1</sup> 张飞萍<sup>2</sup> 李永慈<sup>3</sup> 于振良<sup>1</sup> 杜生明<sup>1</sup>

(1 国家自然科学基金委员会生命科学部,北京 100085; 2 福建农林大学,福州 350002;  
3 北京林业大学,北京 100083)

同行评议是国家自然科学基金项目评审的第一步,也是评审体系中最重要的一环。同行评议是遴选项目的重要依据<sup>[1]</sup>,是公平公正的基础。国家自然科学基金委员会(以下简称自然科学基金委)历来十分重视项目的同行评议工作<sup>[2]</sup>,采取各种措施,如建立高质量的专家库,送审最适宜的同行专家,相近研究项目捆绑式评议等,有效地保证了基金工作的科学民主,公平公正,维护了资助项目的学术声誉。

然而,应当看到,在同行评议工作中还有一些不尽人意的地方,少数专家由于知识面和学识有限,或受社会上讲关系送人情等不良风气影响,其评议意见一定程度地偏离了客观公正,使同行评议工作受到干扰。如何了解同行专家的评议状况,鉴别出不称职的评议专家,优化评议专家队伍,是摆在自然科学基金委项目管理人员面前一项十分重要的工作。诚然,由于同行专家的分散性和不确定性,其研究领域的局限性,学术判断的主观性,加上评议工作受各种外界因素干扰,专家的评议结果成为一个复杂过程的产物,根据评议结果对专家进行评估也就变得比较困难。尽管如此,自然科学基金委仍然重视开展这方面的研究工作,目前在自然科学基金委科研项目网络信息系统(简称ISIS)中特别设置了专家评议状况查询功能,可以方便地获得专家的评议信息,项目管理人员可以根据这些信息了解专家的评议状况,如评议意见是否详细具体,评议指标运用是否合理等。然而,在实际评估中往往发现,一个专家在多个项目评审中对某些项目、某些指标的判断是比较恰当的,而有些是不合适的,根据简单的查询结果很难准确地衡量专家的评议状况。

本文依据基金项目分领域和5位同行专家捆绑式送审的评议方式,利用现有评议系统中专家评议信息<sup>[3]</sup>,采用数学分析的方法,探索设计了一种专家

评估方法。该方法是一种探索,希望通过本文抛砖引玉,不断改进和完善同行专家评估工作。

## 1 评估方法的设计

现行国家自然科学基金项目评审的基本程序是:按研究领域将项目分组,每个项目组选择6—7位同行专家送审评议,同一单位回避,确保每个项目有5位专家评议。评议专家对每个项目的5项指标,即科学意义或应用前景、学术思想的创新性、研究内容、研究方案和课题组的研究能力进行独立判断,并给出由A到D的评价等级、综合评价等级和资助建议,在评审意见栏内详细叙述评价依据。自然科学基金委评审系统将根据各指标的权重和评价等级,计算各项目的综合得分。项目管理人员根据综合得分,分析专家意见和项目内容,决定是否推荐该项目进入学科评审会进行讨论。

根据以上评审程序,依照趋同性原则,我们设计了一种同行评议专家的评估方法。趋同性原则是指,将专家对项目的评审结果与本项目的平均评议结果的符合度作为评价专家的依据,符合度越高专家得分越高,反之专家得分越低。其基本思路是:提取专家评议信息,将定性指标数量化,通过与另外4位专家评议状况比较将量化指标标准化,利用数学方法对标准化数据进行分析,确定专家评估指标,通过评估指标对专家进行排序,筛选出重点评估候选专家,认真分析这些专家评议信息,找出这些专家评议中存在的问题,确定不称职的评议专家。具体过程是:提取专家评议信息→评议指标的量化→量化指标的标准化→获得标准化指标离均差→获得项目各评价指标的综合离均差→获得专家所有评审项目的平均离均差(即综合评估指数)→根据综合评估指数对专家进行排序→获得重点评估专家名单→综

本文于2005年7月22日收到。

合分析这些专家评议信息→发现评议问题→确定较差评议专家的名单。下面详细介绍各个评估步骤。

### 1.1 获取专家的评议信息

在 ISIS 评审系统中,通过输入专家姓名和项目分组标志,获取同行专家的评议信息,整理专家评议指标和评议信息,供以下分析。

### 1.2 评审指标的量化

按照现行的基金项目评审体系对评议指标进行赋值和量化。专家对一个项目有 5 个评价指标,分别是科学意义或应用前景、学术思想的创新性、研究内容、研究方案和课题组的研究能力,上述指标分别被赋予 20、35、15、15 和 15 权重值。专家对每个指标有四个评价等级,分别是 A、B、C 和 D,四个等级分别对应于 1.0、0.8、0.4 和 0.1 四个系数,系数与权重值的乘积即为项目该指标的得分。评审指标的量化用下列公式表示:

$$X'_{ijk} = E_k N_{ijk}$$

$i$  表示专家,  $j$  代表项目,  $k$  为评价指标。  $E_k$  为第  $k$  个指标的权重值,  $N_{ijk}$  为第  $i$  个专家对第  $j$  个项目的第  $k$  个指标指定的等级所对应的系数,  $X'_{ijk}$  为第  $i$  专家对第  $j$  项目的第  $k$  项指标的评议指标的量化值。

### 1.3 评审指标的标准化

由于每位专家评议的项目不尽相同,评议的项目数也有差异,要使专家之间的评议具有可比性,我们对专家的评议结果进行标准化。数学上进行标准化的方法很多,我们在此进行了多种尝试,经过分析和比较,本着科学合理又简便易行的原则,确定以下的标准化方法:

$$X_{ijk} = \frac{X'_{ijk}}{\delta_{jk}}$$

式中  $X_{ijk}$  为第  $i$  位专家对  $j$  个项目第  $k$  项指标评价结果的标准化数据,  $\delta_{jk}$  为第  $j$  个项目的所有评审专家对此项目的第  $k$  项指标评价结果的标准差。

### 1.4 评价指标的标准离均差

以下式计算离均差:

$$Q_{ijk} = X_{ijk} - \bar{X}_{jk}$$

式中  $Q_{ijk}$  为第  $i$  位专家对  $j$  个项目第  $k$  项指标评价结果的标准化离均差,  $\bar{X}_{jk}$  为第  $j$  个项目的所有专家对此项目第  $k$  项指标评价结果的标准化数据的平均值。  $Q_{ijk}$  衡量第  $i$  专家对  $j$  项目  $k$  指标的判定与本项目其他专家的差异和共识性,  $Q_{ijk}$  的绝对值越大,说明专家对  $k$  指标的判断与其他专家的共识性

越小,存在问题的可能性越大,反之,说明共识性越大。

### 1.5 专家评议项目的综合离均差

将 5 个单项指标的标准离均差进行综合,获得专家对该项目评议的综合离均差。由于各个单项指标在具体评价中所占的权重不同,因此在计算中引入单项指标权重系数  $r_k$ 。具体如下:

$$R_{ij} = \sum r_k |Q_{ijk}|$$

式中  $R_{ij}$  为第  $i$  位专家对  $j$  个项目评价的综合离均差;  $r_k$  为第  $k$  个单项指标的权重,当  $k=1$  时,  $r=0.2$ ; 当  $k=2$  时,  $r=0.35$ ; 当  $k=3,4,5$  时,  $r=0.15$ 。显然,  $R_{ij}$  在一定程度上反映了  $i$  专家  $j$  项目评价与其他专家判定的共识度,  $R_{ij}$  越大,共识性越大,  $R_{ij}$  越小,共识性越大,判断越趋于合理。

### 1.6 专家综合评价指数

专家综合评价指数采用下列公式计算:

$$S_i = \bar{R}_i$$

即某一专家对所有评审项目的综合离均差的平均值表示。式中  $S_i$  为第  $i$  位专家的年度综合评价指数,  $\bar{R}_i$  为  $i$  专家当年所有评审项目的综合离均差的平均值。显然,  $S_i$  在一定程度上反映了  $i$  专家当年评审项目的总体质量。  $S_i$  值越大,说明该专家与其他专家的共识度越差,评审质量越差;反之,共识性越高,质量越好。

### 1.7 分析和验证

根据  $S_i$  值对专家进行排序。按照 10%—20% 的比例,筛选出  $S_i$  值较大的专家作为重点评估对象。并在以下几方面分析专家评议中的问题。

$$(1) \text{ 同意资助率} = \frac{\text{同意资助项目总数}}{\text{评议总项数}} \times 100\%$$

该指标在一定程度上反映了专家“评”的作用和掌握评议尺度的能力,资助率极高或极低均说明该专家没有认真比较项目间的优劣。

$$(2) \text{ 优先资助项目的受资助率} = \frac{\text{优先资助项目中实际受资助项目数}}{\text{评议优先资助项目总数}} \times 100\%$$

该指标在一定程度上反映了专家的学术判断力。

$$(3) \text{ 同意资助项目的受资助率} = \frac{\text{同意资助项目中实际受资助数}}{\text{同意资助项目总数}} \times 100\%$$

该指标在一定程度上反映了专家的学术判断力和客观公正性。

$$(4) \text{ 不同意资助项目的受资助率} = \frac{\text{不同意资助项目中实际受资助项目数}}{\text{不同意资助项目总数}} \times 100\%$$

标从另一个方面反映了专家的学术判断力和客观公正性。该指标越高,说明专家越有可能对项目存在主观偏见或学术判断力有问题。

(5) 全面分析专家的评审材料,查看具体的评议意见,查找和分析评议中的问题。如评议的具体和客观性,对哪些指标做出不合适的评价、评价结果偏高还是偏低等。

一旦发现重点评估对象在上述方面的问题突

出,即可确定该专家为不称职的评议专家。

## 2 实例分析

采用上述方法对2004年参加自然科学基金委生命科学部林学学科项目评审的152位专家进行了评估。依据综合评价指数 $S_i$ 对专家进行排序,选取 $S_i$ 排名前15位和后15位专家进行了分析和验证,结果列于表1。

表1 2004年度林学学科同行评议专家反评估结果

专家编号	$S_i$	$S_i$ 值前 15 位				专家编号	$S_i$	$S_i$ 值后 15 位			
		同意资助率%	优先资助项目受资助率%	建议资助项目受资助率%	建议不资助项目受资助率%			同意资助率%	优先资助项目受资助率%	同意资助率%	不同意资助项目受资助率%
F1	1.0746	87.50	27.27	28.57	0.00	L1	0.6383	42.86	100.00	66.67	0.00
F2	0.9499	22.22	0.00	0.00	42.86	L2	0.6325	47.06	—	37.50	0.00
F3	0.9421	30.00	0.00	0.00	14.29	L3	0.6288	58.82	—	20.00	0.00
F4	0.91	42.86	0.00	0.00	25.00	L4	0.6151	66.67	66.67	62.50	0.00
F5	0.8962	44.44	0.00	37.50	10.00	L5	0.5957	60.00	—	50.00	0.00
F6	0.8891	66.67	0.00	30.00	0.00	L6	0.589	29.41	50.00	40.00	8.33
F7	0.8787	100.00	100.00	50.00	—	L7	0.5885	70.00	0.00	28.57	0.00
F8	0.8618	45.45	—	0.00	16.67	L8	0.5853	28.57	—	50.00	10.00
F9	0.8394	27.27	0.00	0.00	25.00	L9	0.582	64.71	75.00	45.45	0.00
F10	0.8332	88.89	75.00	37.50	0.00	L10	0.5672	61.11	100.00	36.36	0.00
F11	0.817	69.23	—	22.22	25.00	L11	0.5578	35.71	—	40.00	0.00
F12	0.8131	16.67	—	0.00	6.67	L12	0.5575	23.08	—	33.33	0.00
F13	0.8109	66.67	33.33	30.00	0.00	L13	0.5487	60.00	100.00	66.67	0.00
F14	0.8057	77.78	50.00	42.86	0.00	L14	0.5481	40.00	50.00	37.50	0.00
F15	0.8029	45.00	33.33	33.33	18.18	L15	0.5131	38.46	—	20.00	0.00

总体上看,表1中 $S_i$ 值前15位专家,同意资助项目受资助率明显低于 $S_i$ 值后15位专家,而不同意资助项目受资助率相反,说明 $S_i$ 值排序反映出了专家的评议质量。在 $S_i$ 值前15位专家中,有5位专家的同意资助率接近或超过70%,同意资助率过高;有6位专家建议优先资助项目的受资助率为0;有6位专家同意资助项目的受资助率为0,有8位专家的不同意资助项目受资助率高于10%。而在 $S_i$ 值后15位的专家中,除未推荐优先资助项目的专家外,所有专家的优先资助项目受资助率均高于50%;建议资助项目受资助率均大于0;有13位专家的不同意资助项目受资助率为0。

分析 $S_i$ 值前15位专家的评议意见发现以下几方面问题:(1)项目评价不够认真,没有很好地比较不同项目间的优劣,只是简单地否定或肯定,如F1、F6、F7、F10、F11、F13和F14号同意资助率过高;(2)项目评价不恰当,如F2—6号和F9号专家,建议优先资助的项目没有1项得到资助,这些专家学术判断力和客观公正性值得怀疑;(3)学术判断力

或公正性有问题,如F2、F3、F4、F8、F9和F12号专家,所有建议资助的项目都没能得到资助,而一些建议不资助的项目最终获得了资助,表明这些专家把握相关研究领域的能力不够,或者未能客观公正地进行项目评审。

## 3 讨论

经由评议指标排序,确定重点评估对象,再结合评议意见进行综合分析,把客观指标与综合分析结合在一起,是一种较为客观、科学、综合和有效的专家评估方法。(1)客观性表现在,将主观评价指标量化、标准化,通过数学分析的方法对指标进行排序,由排序来确定重点评估专家,这一过程完全是数学过程,具有客观和严密性;(2)科学性表现在,除了数学过程的科学性之外,进行了以项目为单位的评议指标标准化,使得各项目评议结果的方差统一,获得的标准化评议指标不但在组内可以比较,而且组间、学科甚至学部内都可以进行比较;(3)综合性表现在,以单个评议指标为出发点,再将多个评价指

标和多个项目评价进行综合,获得专家的综合评价指数,反映了专家整体评审状况;(4)有效性表现为:采用上述方法,对2004年参加林学学科项目评审的152位专家进行了评估,筛选出20位专家进行重点评估,发现15位专家的评议存在较为突出问题,同样确认了20位优秀评议专家,基本上能够反映出专家整体评议状况,具有较高的准确率和可行性。上述专家评估工作为我们优化专家库,进一步提高同行评议质量奠定了良好的基础。我们也曾设想,在不断完善的基础上,将评估结果以一定方式反馈给评议专家,以便让专家更好地了解自己的评议工作,自我约束,不断提高评议质量,实现更好的专家评估效果。

所设计的专家评估方法的另一个特点是操作简便,具有较强的可行性。从整个评估程序看,计算过程较为简单,应用一般的电子表格如Forbase、Foxpro数据库语言或Excel电子表格等均能很容易地完成所有计算过程,操作容易、工作量少、准确性较高,可以在实践中应用。通过综合评价指数( $S_i$ )进行排序,准确有效地筛选出有问题的评议专家,然后对这些专家做进一步的分析,简化了评估过程,提高了准确性,减少了工作量。

尽管本文的专家评估方法具有许多优点,但也表现出一些不足。 $S_i$ 值并不能保证评估结果完全正确。从表1可以看出,23号专家的 $S_i$ 值较小,但其不同意资助项目受资助率仍然较高,同意资助项目受资助率较低,可见该专家在评审过程中也存在

一些问题。出现这种情况与基金评审过程的复杂性和基金项目的多样性有关,主要是:(1)专家评议是定性的,而 $S_i$ 值是定量的,以定性的意见验证定量的结果必然有所差异;(2)同行专家评议只是基金项目评审众多程序中的一个环节,而基金项目的最终结果还受到评审会、申请人完成基金情况等多种因素影响,因此 $S_i$ 值的大小并不能全面体现基金项目受资助的情况;(3)基金项目类别较多,申请项目资助与否受领域影响较大,在竞争激烈的领域,评价较好的项目不一定全部获得资助,而另外一些领域,有些评价稍好的项目即可获得资助,这样直接导致评估结果与实际资助出现偏差;(4)专家评议的项目数在一定程度上影响评估结果。专家评议项目数过少,评议差距往往会比较突出。综上所述,根据 $S_i$ 值,并不能完全正确判断专家的评价状况,只能作为初步的筛选指标,对专家评议状况的最后判定还要研究和专家的具体评议状况。

正如题目所言,本文仅是一个初步探讨,一定有许多不足,供同仁批评指正。

#### 参 考 文 献

- [1] 国家自然科学基金委计划局. 国家自然科学基金项目管理百问. 北京:原子能出版社, 33—42.
- [2] 朱作言. 同行评议与科学自主性. 中国科学基金, 2004, 18(5): 257—260.
- [3] 国家自然科学基金委. 自然科学基金项目管理信息系统. <http://www.nsf.gov.cn>.

### A METHOD TO SCORE THE REVIEW STATUES OF PEER CORRESPONDENTS AT NSFC

Gu Ruisheng<sup>1</sup> Zhang Feiping<sup>2</sup> Li Yongci<sup>3</sup> Yu Zhenliang<sup>1</sup> Du Shengming<sup>1</sup>

(1 Department of Life Sciences, NSFC, Beijing 100085; 2 Fujian Agriculture and Forestry University, Fuzhou 350002;

3 Beijing Forestry University, Beijing 100083)