DOI:10.16262/j.cnki.1000-8217.2003.04.017

·基金纵横·

国家自然科学基金面上项目通讯评议结果的公平化处理

刘 克 王成红 何 杰 宋 苏

(国家自然科学基金委员会信息科学部,北京 100085)

本文对影响项目评审公平性的几种因素进行了分析,用量化的方法对由于专家熟悉程度的不同、评价标准的高低和评价结果偏离度的差异而造成的不公平性进行了调整,给出了统一的计算公式。这种处理方法以统计数据为基础,不带有项目管理人员的主观判断。文中举例验证了这种公平化处理的合理性,同时指出了管理人员在组织项目评审时应注意的问题。

引 言

国家自然科学基金面上项目(以下简称"面上项目")是科学家们自由选题的项目,研究范围广,申请数量大,参评专家多,如何营造一个公平竞争、合理评价的氛围是个值得深入研究的问题。以信息科学部自动化学科为例,2002年收到面上项目申请书550份,涉及6个二级研究领域和49个三级研究方向,每个三级研究方向中的选题又因理论、方法和应用背景的不同而分为若干更细的分支,有近400位同行专家参与了项目的通讯评议。项目管理人员(项目主任)在收到专家的《评议意见表》后,一般以学科为单位对数百个项目进行统一排序和分类,这种"分散评审、统一排序"的遴选方式可能存在以下问题:

(1)不同专家的评价标准(心理水准)存在明显差异,有的专家偏于严格,评价等级普遍偏低;而有的专家则比较宽松,评价等级大多偏高。

(2)基金项目多属前沿领域,鉴别工作有较大难度;许多项目还涉及学科交叉,需要不同学术背景的专家进行评审。有些专家对研究内容不是非常熟悉,反映到《评议意见表》中,这些专家对研究领域熟悉程度的自我评估是"较熟悉"或"部分熟悉"。

(3)由于个人、研究组或单位之间的一些非学术方面的原因,个别评议人对项目评价不够客观,其评议结果与其他评议专家的评议结果和最终表决结果有较大偏离。

这些因素可能使某些项目申请受到不公正待遇,有必要进行一定程度的公平化处理。

1 公平化处理过程

每份面上项目申请书发给5位同行专家进行通讯评议,收回3份《评议意见表》即为有效评议,一般收回4—5份。每份《评议意见表》中包含三项内容:

(1)专家对项目涉及领域的熟悉程度(以下简称熟悉程度):A(熟悉)、B(较熟悉)、C(部分熟悉)。

(2)专家对项目申请的综合评价等级:A(特优)、B(优)、C(良)、D(中)、E(差)。

(3)以文字表述的具体评价意见。

一个项目的综合评价包括同行专家的评审意见、项目主任的分析判断和基金政策的各种奖惩因素,对专家意见的分析处理是综合评估的一个重要组成部分,应力求客观公正。有些项目主任根据专家给出的评价等级进行定量计算和排序,朝定量化方向走出了一步。更进一步看,《评议意见表》中还有熟悉程度信息可以利用,而专家潜在的评审特点也对公平性存在影响,也应予以考虑。本文依据《评议意见表》中第(1)、(2)项的内容,在分级量化的前题下,建立一种基于统计数据的公平化处理方法。第(3)项内容包含许多重要信息,直接影响项目主任的主观判断,对这一部分信息的处理留待项目综合评价时进行。

1.1 评价等级和熟悉程度的量化

《评议意见表》中"综合评价等级"一栏有5个等

本文于2003年4月17日收到.

级，可以用 5 分制计算专家评分(G)，这是一个约定俗成的量化方法，如表 1 所示。有的专家曾给出过 B－或 C＋这种更细的评级，简单用 B 或 C 代替不够准确，此时可考虑进行插值。需要注意的是，B－和 C＋在感觉上是有区别的，把它们定位在相邻两级之间的 1/3 处和 2/3 处比较合适。

专家的熟悉程度按表 2 进行量化，熟悉程度系数 F 的取值也是根据经验确定的，没有绝对的衡量标准，但它对所有项目都是一致的，从这一点来看它是公平的。

表 1 评价等级的量化

评价等级	A	A－	B＋	B	B－	C＋	C	C－	D＋	D	D－	E＋	E
G 的分值	5.0	4.7	4.3	4.0	3.7	3.3	3.0	2.7	2.3	2.0	1.7	1.3	1.0

表 2 熟悉程度的量化

熟悉程度	熟悉	较熟悉	部分熟悉
F 的取值	1.0	0.8	0.6

1.2 根据专家熟悉程度进行的调整

一个必须承认的事实是，熟悉研究内容的专家认定的优和较熟悉的专家认定的优，其份量是不同的，“部分熟悉”的专家给出的过高或过低的评价应按照其熟悉程度进行折扣。

这里首先有一个选择基准值 G_0 的问题，由于专家评分 G 是从 1 开始取值，按照现有的评价参考标准，1 分和 2 分主要是负面意见，3 分以上才是正面意见，因此 G_0 不能直接取 0，而应该取在 2 和 3 之间。可以简单取 $G_0=2.5$，即 2 分(弱否定“暂不考虑资助”)和 3 分(弱肯定“可以考虑资助”)的中值。

启用熟悉程度系数之后，专家评分 G 的一次调整值 G^1 由下式给出：

$$G^1 = (G - G_0) \cdot F + G_0 \qquad (1)$$

专家评分 G 与基准值 G_0 的偏差$(G - G_0)$无论为正还是为负，都要依据熟悉程度系数 F 进行调整，然后加上基准值 G_0，还原到[1,5]标度。F 的缺省值为 1，此时不考虑熟悉程度因素，G^1 退化为 G，相当于专家评分被全部采纳。

1.3 根据专家评价水准进行的调整

送给评审专家的材料中都会附上一份评价参考标准，绝大多数评审专家也都知道面上项目的资助比例，但从评价等级的平均值上看，不同专家的评价水准仍然存在明显差异，为保证所有项目申请都能受到相同或相近的“待遇”，有必要进行一定的平衡处理。

具体做法是，将某专家的平均评分与全部专家的平均评分进行比较，得到该专家的评价水准系数 H，用 H 对该专家的评分进行整体升降。一个专家当年评分的平均值，$Ea = \frac{1}{m}\sum_{j=1}^{m}[(G_j - G_0)\cdot F_j + G_0]$，m 是该专家当年评议的项目数。全部专家当年评分的平均值 $E_0 = \frac{1}{n}\sum_{i=1}^{n}[(G_i - G_0)\cdot F_i + G_0]$，n 是当年所有《评议意见表》的总数。H 的计算方法是 $H = \frac{Ea}{E_0}$，从理论上讲，H 的取值范围是开区间(0.2, 5)，但在实际中 H 一般落在[0.75,1.25]区间。

计入熟悉程度和评价水准系数后，得到评分 G 的二次调整值 G^2：

$$G^2 = [(G - G_0)\cdot F + G_0]\cdot\frac{1}{H} \qquad (2)$$

当该专家对所审项目的评价等级整体偏高时，H 大于 1，在(2)式中体现为对其评分向下调整；当该专家的评价等级整体偏低时，H 小于 1，在(2)式中体现为对其评分向上调整。H 的缺省值为 1，此时不考虑专家评价水准的差异。

需要说明的是，Ea 偏离 E_0 可能由三个原因所致，H 的作用也不尽相同：

(1)专家评审的项目数量较少，项目水平有一定的偶然性。此时 H 的使用会把该专家的评分拉向均值 E_0，这种调整显然不够合理。为了降低这种不合理性，一是在送审阶段适量增加专家评审项目的数量，二是在后期分析阶段，只对评审数量超过一定上限(例如 $m \geqslant 10$)的专家启用 H 系数进行分值调整。m 越大，H 的作用越合理。

(2)专家的心理水准偏高或偏低。此时 H 的作用是客观而公正的。

(3)不同学科分支的研究水平存在差异。这种情况下专家给一部分项目打分偏高或偏低是符合实际的，利用 H 进行平衡处理有失客观。但此时 H 的调整功能可以看作是学科分支间的一种平衡作用，即对过热分支进行适量限制和对弱小分支进行适量扶持，从管理的角度来看，这种做法对一个学科的均衡发展不无益处。另外，从实际情况来看，当专家评审项目较多时，被评项目一般不会集中在某一狭小的学科分支内，而是覆盖几个相近的学科分支，打分偏高或偏低可以认为主要是由于心理水准的差异而

造成的。

1.4　根据反向评估情况进行的调整

以面上项目最终审定结果为依据,对专家评议意见的准确度进行反向测算,以判定其学术鉴别能力和评价的客观性,不仅有利于新一轮项目评审的公平性,也有助于动态调整评议专家队伍。

设T是某位专家近年来评议的项目总数,Ts是其评判结果(同意或不同意资助)与最终集体表决结果(给予或不给予资助)相同的项目数,则该专家的反向评估调整系数可用 $R=\frac{Ts}{T}$ 简单计算。R的取值在[0,1]区间,评判的准确度越高,R值就越接近1;若该专家的评判结果总是与最终结果相差很大,则R值较低。R的作用与熟悉程度系数F相似,都是对专家意见的份量进行调整,即对偏差值$(G-G_0)$进行修正。

计入以上三项调整系数后,得到G的三次调整值 G^3:

$$G^3=[(G-G_0)\cdot F\cdot R+G_0]\cdot\frac{1}{H} \qquad (3)$$

考虑到评审专家同意(或反对)资助一个项目的程度是有区别的,还可以利用这种程度上的偏差对R进行更精确的计算,如 $R=1-\frac{1}{T}\sum_{i=1}^{T_s}\left|\frac{1}{5}G_i-V_i\right|$,其中,$V_i$是一个项目的综合得票率(含通讯评议专家的资助态度和学科评审会专家的投票情况),是该项目水平的比较客观的度量,$0\leqslant V_i\leqslant 1$。关于这一算法的详细分析另文再叙。

反向评估工作应该在历史数据比较完整的情况下进行,当历史数据不足时,应停用该项功能,此时R取缺省值1。

1.5　项目的平均分数

一个项目的平均评价分值S由下式给出:

$$S=\frac{1}{n}\sum_{i=1}^{n}[(G_i-G_0)\cdot F_i\cdot R_i+G_0]\cdot\frac{1}{H_i} \qquad (4)$$

式中n为该项目的《评议意见表》数量($3\leqslant n\leqslant 5$),G_i、F_i、H_i、R_i是专家Ei的评定分值及各项系数,将一个项目的n个专家评分经处理后求平均值,即得到S。

在截止期限内返回的《评议意见表》可能少于5份,而一份评议表的最低评分为1,少一份会影响到一个项目的总积分,因此按照平均分而不是总积分进行排序是比较合理的。但是按照S进行排序也存在问题,同时具有5份评议表的项目之间进行比较才是公平的,3—4份评议表的平均值不一定能代表缺少的那1—2个专家评级,尽管从数学意义上说这种估算是最接近的,但其中仍然含有不确定性。从实际情况来看,专家意见经常有分歧,缺少的那1—2个专家评级很可能高于或低于平均水平。

2　实例分析

以下结合一个例子予以说明。有10个项目P01—P10,分别送交8位专家E1—E8进行评议,共返回50份《评议意见表》,所得结果和简单排序如表3所示。

表3　项目评价与简单排序

项目编号	特优(份)	优(份)	良(份)	中(份)	差(份)	回函份数	简单计分	平均分值	简单排序
P01	1	2	2			5	19	3.80	1
P02	1	1	2	1		5	17	3.40	2
P03		2	2	1		5	16	3.20	3
P04		2	1	2		5	15	3.00	4
P05		1	3	1		5	15	3.00	4
P06		1	2	2		5	14	2.80	6
P07		1	1	2	1	5	12	2.40	7
P08			3	1	1	5	12	2.40	7
P09			2	2	1	5	11	2.20	9
P10			1	3	1	5	10	2.00	10

专家评分和熟悉程度的详细情况如表4所示。可以看出,任意两个项目都有两个以上的相同评议人,评审项目最多的专家E5共审了8份,最少的E7审了4份,其余的专家审了5—7份。72%的《评议意见表》中的专家熟悉程度系数为1(熟悉),其余都是0.8(较熟悉)。这种专家指派情况在实际操作中是合理的。

鉴于历史数据不足,本例中未启用反向评估功能,即8位专家的反向评估系数R均取1,此时 $G^3=G^2$。依照(4)式进行精确计算和排序的结果如表4所示。其中,So表示项目的精确排序,Ch为处理后项目排序的升降情况。每位专家评分的平均值Ea及其评价水准系数H列于表5。

经过公平化处理后,项目排序发生了一些变化,最重要的变化是P02由第2位下降到第3位,而P03由第3位上升到第2位。对于面上项目20%左右的资助率来说,P02和P03的换位意味着P03得到资助的机会大增而P02得到资助的机会锐减。当有多个项目聚集在资助边缘一带,从直观上难以区分时,精确排序的作用会更加明显。

现从直观上简单分析一下这种变化的合理性。有两位专家(E2和E6)给了P02较高评价(5分和4分),但他们的自我评估都是"较熟悉"而不是"熟

悉",并且打分又整体偏高,经过折扣后,P02 的分数有所下降是符合情理的,而这一分数的下降直接导致了 P02 排名的下降。

表 4 专家对项目的评定情况及精确排序

	E1			E2			E3			E4			E5			E6			E7			E8			S	So	Ch
	G	F	G^3	G	F	G^3	G	F	G^3	G	F	G^3	G	F	G^3	G	F	G^3	G	F	G^3	G	F	G^3			
P01	3	1.0	2.7	5	0.8	3.7							3	1.0	3.9	4	1.0	3.4				4	0.8	3.8	3.520	1	→
P02				5	0.8	3.7	3	1.0	2.9				3	1.0	3.9	4	0.8	3.2	2	1.0	2.5				3.245	3	↓
P03	4	1.0	3.6				4	1.0	3.9	3	1.0	3.1	3	1.0	3.9				2	1.0	2.5				3.416	2	↑
P04	4	0.8	3.3				4	0.8	3.6	3	1.0	3.1	2	1.0	2.6							2	1.0	2.1	2.960	4	→
P05	4	0.8	3.3	3	1.0	2.5				3	0.8	3.0				2	1.0	1.7	3	0.8	3.6				2.838	5	↓
P06				3	1.0	2.5	3	0.8	2.8	2	1.0	2.1	2	1.0	2.6							4	1.0	4.1	2.836	6	→
P07				3	1.0	2.5	2	1.0	2.0	2	1.0	2.1	1	1.0	1.3	4	0.8	3.2							2.203	9	↓
P08	3	1.0	2.7	2	1.0	1.7				3	0.8	3.0	1	1.0	1.3							3	1.0	3.1	2.365	7	→
P09	3	1.0	2.7				2	0.8	2.1				2	1.0	2.6	3	1.0	2.6				1	0.8	1.3	2.258	8	↑
P10	1	1.0	0.9				2	1.0	2.0							3	1.0	2.6	2	1.0	2.5	2	1.0	2.1	1.998	10	→

表 5 专家评分的平均值及其评价水准系数

	E1	E2	E3	E4	E5	E6	E7	E8	总平均
Ea	3.06	3.33	2.81	2.63	2.13	3.23	2.23	2.67	$E_0 = 2.76$
H	1.11	1.21	1.02	0.95	0.77	1.17	0.80	0.96	

3 结 语

通过以上分析计算可以看出,这种基于量化统计的公平化处理具有明显的合理性,在目前项目评审过程中还存在一些不平衡、不客观因素的情况下,使用这种处理方法对贯彻"公平竞争、择优支持"的方针可望起到积极作用。

对专家意见的公平化处理只是项目综合评价工作的一部分内容,更全面的评价指标还应该包括项目主任的分析判断和政策调整因素。需要说明的是,项目经过综合评价后的排序和分类不是最终结果,而是终审专家们进行复议和投票时的一个参考,"依靠专家、发扬民主"是基金项目评审的一贯原则。

本文提供的处理方法中还有一些问题值得深入研究,例如:评价等级和熟悉程度的量化标准仍然存在人为因素,级差的大小直接关系到调整作用的强弱;以统计数据为基础的 H 和 R 系数也倾向于钝化专家们的个性评价和抹平学科分支间的水平差异;H 和 R 在启用前后取值是不连续的,由此会带来计算公式上的二次不公平问题。任何处理都会带有管理者的主观色彩,不处理就是最好的处理,而要使处理作用尽可能弱化,就要尽力使各项调整系数接近 1,这从管理的角度给项目评审的组织者提出了一些建议:

(1)同类研究项目尽量送相同的一组专家,以营造相同的评价环境。

(2)尽量选择熟悉研究领域、评价公正且心理水准与参考标准一致的专家。

(3)让每位专家评审足够数量的申请,以便横向比较和后期统计分析。

(4)尽量在规定的期限内收齐 5 份《评议意见表》,去除不确定因素。

(5)对专家的历史评审数据进行必要的记录和分析,动态调整专家队伍。

参 考 文 献

[1] 吴述尧. 同行评议方法论. 科学出版社, 1996.

[2] 商玉生. 十岁的思考. 中国言实出版社, 1997.

[3] 王志强等. 关于完善同行评议制度的若干问题和思考——同行评议调研综述. 中国科学基金, 2002, 16(5):309—313.

[4] 夏传钺. 同行评议专家的经验权重. 中国科学基金管理专刊, 1995.

[5] 杨列勋. 对基金项目同行二次通讯评议的案例分析. 中国科学基金, 2003,17(1):50—53.

REDRESSING THE BALANCE AMONG REVIEWER COMMENTS ON NSFC GENERAL PROJECTS

Liu Ke Wang Chenghong He Jie Song Su

(*Department of Information Sciences, NSFC, Beijing* 100085)