

·基金纵横·

# 对基金项目同行二次通讯评议的案例分析

杨列勋

(国家自然科学基金委员会管理科学部,北京 100085)

## 1 问题提出

目前,同行评议在项目的选择与评价方面得到了广泛的应用。但是,随着同行评议的大量使用、应用对象的多样性和专家网络关系的复杂性,它的不足之处也逐渐暴露出来,其中之一是其公正性、公平性受到一些被评估者的怀疑。他们认为同行评议不能排除人情干扰,结果有一定的随意性,并不能将最优的项目选择出来<sup>[1]</sup>。一个极端的情况是对于同一个项目或论文,如果换了一组同行专家重新评议,即进行同行二次评议,则会得到截然相反的结论。即同行二次评议与初次(一次)评议结果会有相当大的差异。美国国家科学基金会的研究表明<sup>[2]</sup>,有约占30%的项目,如果同行评议人改变,则评议结果也会改变。国家自然科学基金委员会是我国较早运用同行评议方法开展项目评价的机构之一,近几年国家自然科学基金委员会每年收到3万余项申请,每年有1万余名专家参加同行通讯评议,800余名专家参加学科评审会。基金项目的同行评议过程中是否存在评估结果的不确定性?如果有,产生变异的项目有什么特点,如何分布?会对最后的结果产生何种程度的影响?这些都是大家十分关心、非常想了解的问题。针对上述问题,笔者专门以一组基金项目为案例进行试验,将同行二次评议与初次评议结果进行比较,统计并分析产生不同评价结果的原因,并就发现的问题提出了相关的改进建议与意见。

## 2 案例研究

国家自然科学基金委员会的同行评议过程包括同行通讯评议和学科评审组会议评审。考虑到学科评审组评议过程的不可重复性和专家评审会的严肃性、权威性和唯一性,本案例只对同行通讯评议进行了二次评议的试验。

选择同行通讯评议进行案例试点主要考虑到:(1)通讯评议宜于操作和过程控制;(2)通讯评议的重要性。通讯评议虽然不能决定可以资助哪些项目,但它可以决定哪些项目不能得到资助。由于申请项目数量远大于资助立项数,受资助比率的限制,有70%—80%的项目在同行通讯评议过程中遭到淘汰。因此,同行通讯评议在国家自然科学基金委员会的评审过程中占有重要地位,它是获得资助项目必须经过的一道程序。

我们设计对一组申请项目作同行二次通讯评议,该组项目经初次评议后,再选择多组同行专家进行二次通讯评议。

### 2.1 试验的基本情况与结果

试验范围:某一组自由申请项目共10项,这些项目均属于同一研究领域,所选择申报的学科代码和关键词基本一致。

试验方式:先将10份项目申请送一组专家(5位)进行初次同行通讯评议,在收到5位专家的评议意见之后,再选择A、B、C三组同行专家进行二次通讯评议。三组专家情况分别如下:

A组:为初次评议的5位专家,告知初次评议的各个项目的具体意见,请他们重新评议。

B组:重新选择5位专家,告知已经初次评议并提供针对每一项目的原先5位专家的具体意见作参考,请重新评议。

C组:不告知初次评议的结果,请5位专家进行评议。

所选择的评议专家对该研究领域熟悉,三组专家的成员不重复。专家在定性判断和评估的基础上,必须对项目的总体情况给出特优、优、良、中、差的定量评价,对其中某一具体项目内容不熟悉的专家评议意见则视为无效。

共有13位专家参加了同行通讯评议的试验工

本文于2002年11月20日收到。

作,其中 A 组专家和初次评议专家是同一组人员,另外 2 组 8 个专家则是重新选择的专家。A 组的 5 位专家全部反馈意见,B 组和 C 组均返回 4 位专家的意见,且每一项目的每一组评议结果均有 3 份以上的有效评价意见。

我们对专家反馈的评议结果进行处理。首先对

项目进行编号,把专家对某一具体项目的评议结果按组为单位将意见集中起来。A、B、C、D、E 分别代表评价结果为特优、优、良、中、差的情况,各计为 5、4、3、2、1 分。对定量评价结果按平均得分多少进行排列,得分最高的项目评价最好,得分最低的项目评价最低,具体得分排列如表 1 所示。

表 1 各组专家评价结果对照表

项目序号	一次评议结果	二次评议结果			平均*	一次评议主要意见
		A 组	B 组	C 组		
1	2A2B1D 4	2A2B1C 4.2	2B2C 3.5	2A2B 4.5	4.1	总体评价高,有创新,熟悉专家评价很高、较熟悉专家评价较低
2	2B3C 3.4	2B3C 3.4	4C 3	2C2D 2.5	3	非理论性研究,创新不明显
3	1A1B2C1D 3.4	2B3C 3.4	1A1B2C 3.75	1A2B1D 3.75	3.62	对国外研究趋势了解不深,创新性不突出
4	3B2D 3.2	3B2D 3.2	2B2C 3.5	1B3D 2.5	3.08	项目偏伦理学、社会学;申请者基础较弱
5	2B1C2D 3	1B2C2D 2.8	3D** 2	1C2D1E 2	2.3	立论不充分,内容过大
6	2B2D1E 2.6	2B3D 2.8	2B1C** 3.67	2B2D 3	3.1	有意义,申请者人积累少,项目设计粗糙
7	1B1C3D 2.6	1B4D 2.4	4D 2	1B1C2D 2.75	2.58	基础不够、对现状不了解
8	2C3D 2.4	2C3D 2.4	1B2C1D 3	3C1E 2.5	2.62	有意义,立论、内容设计及对国外进展尚需进一步了解
9	1C4D 2.2	5D 2	4D 2	1A1B2D 3.25	2.38	前期积累不够,立论不充分
10	3D2E 1.6	3D2E 1.6	1C3D 2.25	1B1D2E 2	1.92	立论不充分,创新不足

\* 表中平均为 ABC 三组 13 人的平均得分; \*\* 只有 3 个评价意见的,表明有一份为无效意见。

表 1 反映出专家有自己独立的学术思想和观点,二次评价不易受前次评议的影响。一次评议选择的 5 位专家也就是 A 组专家基本坚持原先的意见,A 组评价结果及各个项目的排序与一次评议相比没有变化。B 组专家在已知一次评议专家意见的基础上,对各项目的评价与一次评议的意见差异很大。C 组专家在没有告知一次评议结果的前提下,其提出的评议意见与一次评议意见有一定的相似性,差异相对较小。对表 1 中 4 组评议结果按项目得分高低进行排序,顺序如表 2 所示。

如果将 10 个项目分成优良、中等、差三个小集合,每一个集合中平均含三个项目,则得到表 3 的等级分类表。

二次评议后,A 组优良、中、差的三组集合中成员与一次评价完全一样。而 B、C 两组专家的评议结果则与一次评议结果有一定的差异,拟合率(一致率)只有 50% 左右。从排序统计看,A 组的专家还

是坚持原先的意见,排序顺序并没有发生变化。B 组专家给出的结果与 A 组及参照组的差异比较大,前三名中只有 3 是一致的。C 组与 A 组及参照组的排序顺序也有一定差异,前三名中有二名 1、3 是一致的。

表 2 项目得分排序表

项目序号	一次评议排序	二次评议排序			平均得分
		A 组	B 组	C 组	
1	1	1	3	1	1
2	2	2	6	3	3
3	3	3	1	9	6
4	4	4	4	6	4
5	5	5	2	7	2
6	6	6	8	2	7
7	7	7	10	4	8
8	8	8	5	8	9
9	9	9	7	5	5
10	10	10	9	10	10

注:一次评议中,2 与 3 并列;二次评议 A 组中,2 与 3 并列,5 与 6 并列,7 与 8 并列;B 组评议中 1 与 4 并列,2 与 8 并列,5、7 与 9 并列;C 组评议 2、4 与 8 并列,5 与 10 并列。

表3 项目等级排序表

项目等级	一次评议结果	二次评议结果			平均
		A组	B组	C组	
优良(前三名)	1,2,3	1,2,3	3,6	1,3,9	1,3,6
中等(中间三名)	4,5,6	4,5,6	1,4,2,8	6,7,2	2,4,7
差(最后四名)	7,8,9,10	7,8,9,10	10,5,7,9	4,8,5,10	8,9,5,10

注: B组中由于1与4, 2与8得分相同, 故一并列入中等。

由于 NSFC 的项目资助数量有限, 只有 20% 左右的项目经同行通讯评议后能够提请评审会讨论, 因此 10 个项目中可以选送 2 个项目。如果送审 2 个项目进入后一轮的评审会, 则根据排序前 2 名的次序可得到: 一次评议中项目 1 和 2、3 送审; 二次评议中各组送审情况为: A 组项目 1 和 2、3; B 组项目 3 和 6; C 组项目 1、3、9 送审。一次评议的选择送审结果与二次评议的选择送审结果并不完全一致。

在实际的评议操作中, 如果换一组专家重新评议项目, 不告知一次评议的结果, 则只出现一次评议和二次评议 C 组的情况, 这两组的送审项目中, 有 1、3 是共同推荐的, 一次评价还推荐了项目 2, 而 C 组则推荐了项目 9。

上述结果表明, 如果换一组专家, 评价结果确实会出现一定程度的变化, 存在一定的偏差。在国家自然科学基金委员会当年度的实际评审过程中, 因为存在与其他不同领域的项目组评议结果的平衡和比较, 该组只有项目 1 提交到评审会上, 且在评审会上通过专家组的评议, 得到资助。

## 2 原因分析

二次同行通讯评议案例试验表明, 对于大多数项目, 评价较高的依然较高, 评价较低的仍然较低, 评价中下游的仍处在中下游。较之排序前面的项目, 处于中间位置的项目排序变动更大。为什么会存在排序不一致的现象, 笔者认为主要有以下原因:

(1) 专家对项目评价标准的把握不一致。国家自然科学基金委员会目前的评价准则主要强调项目的立论依据、研究内容设计、研究方法与技术路线的合理性、前期研究基础及经费预算等方面, 突出研究项目的创新性。各位评议专家对上述几个具体标准的重要性的理解并不完全一致, 如产业部门的专家往往更看重潜在应用前景和重要性, 而大学里的教授更偏向于学术理论和研究创新。这种倾向毫无疑问会影响到具体项目的评估结果。

(2) 专家的学识和研究方向的差异导致对项目

评估结果的不一致。一般认为, 同行相轻, 但在实际的评议过程中, 我们发现专家更容易保护本领域的利益和项目, 对于别的领域的项目往往要求较严, 在小同行和大同行的申请项目中, 更倾向同意小同行的项目。尽管我们选择的都是同一小领域的专家, 但每一专家的研究方向仍存在一定的差异。这种差异会引起对项目评价结果的差异, 这从每位评议专家对每一项目是否熟悉的表态和评议结果的初步分析中可以发现。

(3) 人情关系。C 组评议结果中, 项目 9 在一次评议中排第 9 名但在二次评议 C 组中排第 3 名, 不仅与一次评议组的评议结果出入很大, 与 13 人平均的结果也有很大出入, 有点出人意料。经核实后发现, C 组的 4 位专家中给项目 9 以特优的评价者与申请者的个人关系非常好, 人情关系确实在此影响了评价结果。

## 3 初步结论与有关建议

通过上述案例分析, 我们对同行通讯评议系统得出如下初步结论:

(1) 同行通讯评议系统是公正和有效的。虽然一次评议和二次评议有一定差异, 但它能辨识最好的项目和最差的项目, 通过它可以最优秀的项目选择出来, 同时也能阻止最糟的项目得到资助。由于所有项目均由同行专家评议, 确保项目过程的民主和公平。在项目资助数量有限的前提下, 同行评议所产生的不确定性和随意性对结果的影响并不很大。因为在多个专家群决策的环境中, 一个隐含人情关系的专家意见并不能决定一个项目的命运。

(2) 对于评价排序在中间部分的项目, 一次和二次评议的结果有一定的分歧。如果按 15% 的资助率来选择送审项目, 则依然可以保证排在前 10% 的项目是有效的、优秀的。但是对于非共识的项目来说, 则存在一定的随机性。因为非共识项目的得分排序一般不会出现最前面, 也不会出现在最后面, 大多数出现在中间位置。在资源有限、资助率控制较严的前提下, 这些非共识的项目难于全部得到机

会被提交到评审会,否则会挤掉评价很好排序靠前的项目名额。

(3)同行通讯评议确实存在一定的不确定性,这对参加评审会的专家提出了更高的要求,通过学科评审会的形式再一次进行同行评议将真正具有创新性、研究基础好且有可能取得成果的项目选择出来是评审组的权利,更是责任。

本着保护创新思想、鼓励科学家自由探索的精神,针对同行评议过程及项目评价方面的问题,提出几点个人的想法。

(1)要充分发挥项目管理人员的作用。项目主任、学科处长对同行评价队伍要尽量熟悉和了解,特别是对于同行之间的关系和不同学术流派之间的情况要了解,以避免选择专家时陷入庸俗的人际关系圈,要尽量使这种不正常、非学术的影响因素在同行评议中降到最低。同时,项目管理人员要有辨识非共识项目中真正具有创新性学术思想研究项目的的能力,尽量保证将此类项目提交到评审会,而不是简单地将此类项目的定量得分高低作为送审项目的唯一依据,对基础研究项目的评价,在关注定量的结论性评价结果的同时,还应更多地关注定性的专家评价意见。

(2)评价标准的设计要有导向性。目前的评价标准对申请者的研究基础、学术思想的创新性突出的仍然不够。个人认为,对一个项目的评价,主要关注两个方面:项目的学术创新和申请者本人的条件。其他的如经费预算、预期成果等则是次要的因素。因为基础性研究本身就是一种不确定的探索,很难在项目设计之初就可以预计到今后可能会出现什么样成果。经费预算更是有很大的随意性,完全是为了预算而做预算,不应将这种缺少严谨性和科学性的问题作为一个重要的评价标准。

(3)在条件允许的前提下,尽可能的多选择几位

同行通讯评议专家进行评议。在可能的情况,可以去掉一个最高分和一个最低分,这也是减少人情关系影响的有效方法之一。

(4)扩大资助率,增加送审项目数量。由于基础研究的不确定性和同行评议方法存在一定的随机性,增加资助率,扩大资助面,会使更多有一定创新的、非共识的项目得到资助的可能,并且最终会得到一些意想不到的成果。同行通讯评议的结果表明,处于评价中间部分的项目有些确实具有一定的创新性,这些项目往往有争议,如果把这部分项目选择进来,在可能的情况下加以资助,则与保护创新的学术思想、创造宽松学术环境的指导思想并不相悖。

(5)加强学科评审组专家的择优功能和学术把关的责任。学科评审组的重要功能不仅体现在最后的项目选择决策权,更表现在对科学领域、关键问题、创新思想的把握和判断上,在经费资源有限的前提下,评审组一定要把好关,树立保护创新的评审原则和基调。

(6)利用文献计量方法进行学术水平和研究基础的定量跟踪。看重研究者过去的积累和已有研究基础的学术影响,可以利用一些检索系统如SCI、SSCI、CSCD、CSSCI等的论文发表、引用结果作为评价的辅助的定量依据,增加评价的客观性。

本次案例仅是初步探索,在今后的案例跟踪过程中,我们将增加案例项目数量、关注专家定性评价意见、并实施全过程跟踪试验,以期逐步掌握同行评议过程的内在规律。

#### 参 考 文 献

- [1] 吴述尧. 同行评议方法论. 北京:科学出版社,1996.
- [2] 沈新尹. 同行评议新动态——从美国审计总署的调查报告到国家自然科学基金会的专题讨论. 世界科技研究与发展,1996,(6):73—76.

## SECOND PEER REVIEW OF FOUNDATION PROJECTS——A CASE STUDY

Yang Liexun

(Department of Management Science, NSFC, Beijing 100085)