

·基金纵横·

对同行评议专家评议工作进行评估的一些思考

何杰 王成红 刘克

(国家自然科学基金委员会信息科学部,北京 100085)

国家自然科学基金项目评审工作主要包括三个环节,即同行通讯评议、项目主任综合同行通讯评议意见和学科评审组会议评审,同行通讯评议和学科评审组会议评审都属于同行评议。其中,同行通讯评议所起的作用不仅至关重要,而且也是后两个环节的基础。同行评议是基金项目评审的核心,也是落实国家自然科学基金委员会“依靠专家,发扬民主,择优支持,公正合理”评审原则的关键所在。基金项目和基金评审工作在科技界的良好声誉很大程度上得益于同行评议制度。尽管目前还没有能够取代同行评议的更好办法,但同行评议也有其不足之处,如某些创新性强的项目不易得到较高评价,学科交叉性强的项目不易取得共识等。对此,我们只能采取相应措施,提高同行评议质量,进而充分发挥同行评议的作用。

同行评议活动主要包括被评议对象(项目申请书所含内容)、评议专家和评议标准三个要素。在基金项目评审过程中,当申请书被受理后,评议该申请项目的同行评议专家的知识背景、对评议标准的理解及其价值观就成为决定其评议结果的主要因素。因此,同行评议专家的选择至关重要。有关如何选择同行评议专家的研究成果及相关论述已有不少积累,其中,将不同因素所起作用的大小进行量化并形成选择同行评议专家指标体系的方法也已受到关注^[1-3]。例如,选择同行评议专家可以根据其基本情况、修养和工作业绩三个指标体系来考察。基本情况指标体系包括职称、学历、科研工作、发表文章、所获奖励、人才培养、年龄和性别等众多指标;修养指标体系包括品德修养(包括责任感、科学性、公正性、相容性)、认知素养(包括准确性、敏捷性、独立性)、意志品质(包括果断性、自觉性)、智能修养(包括观察力、思维力、创造力、表达力、信息处理能力)、知识素养(包括学术水平、了解动态深度)等至少 5

类 16 个指标;工作业绩指标体系包括评议项目累计数、离散率、命中率、成功率等指标。由于这些指标体系非常复杂,有些指标难以获得,有些指标的科学性也有待验证,因此在实际操作中,只具有参考作用。

通常我们在初次选择一位专家作同行评议人时,都会对其进行一些预评估,如考虑其对相关领域的熟悉程度、学风是否公正、对科学前沿是否敏感、科研成果、发表论文、职称等等因素。当同行评议专家参与评议工作后,对所选同行评议专家的评议工作进行适当评估,引入部分工作业绩指标,了解专家在评议工作中的表现,有利于今后对同行评议专家的筛选,形成基本评议专家队伍,建立专家库,对项目主任的工作大有益处,在此我们不妨称之为对同行评议专家的反评估。国家自然科学基金委员会信息科学部原半导体学科主任夏传钺研究员曾在此方面做了大量的工作,取得了一些有意义的结果^[4]。近年来,我们仍在坚持这方面的探讨工作并积累了一些经验和体会,下面谈一下我们的体会。

1 对同行通信评议专家的反评估也是对项目主任自身的评估

同行评议中决定性因素虽然是同行通信评议专家,但同行通信评议专家是由项目主任指派的,因此对同行通信评议专家的反评估也是对项目主任自身的评估。比如评议人对所评项目是否熟悉,就需要项目主任对申请书的内容和评议人的工作背景有相当的了解,而这些了解同时还受到项目主任自身知识水平的制约。因此,项目主任必须经常参加相关的学术活动,经常深入到科研第一线进行学习和调查研究,自觉地提升自身的学术水平。项目主任在选择同行通信评议专家时,既要注意积累平时收集到的信息,也要加入自己的判断。评议人的性格各

本文于 2003 年 10 月 15 日收到。

异,有些评议人常称自己对某项目不熟悉或较熟悉,而实际他对该项目相当熟悉;也有个别评议人对什么项目都表示熟悉,其实多数情况下只是略知皮毛,这些情况都需要项目主任深入了解。因此,应该为项目主任提供必要条件,使其有机会、有条件提升自己的学术水平,并尽可能多地与广大科技工作者进行较深入的接触,以便了解到更多有用的信息。

2 掌握评议专家对待评议工作的态度

掌握评议专家对待评议工作的态度,重点考察其认真程度。有些专家对自然科学基金不很了解,对基金项目的评审不热心,这就需要在有机会的时候多介绍基金的有关情况,宣传基金在我国基础研究中的重要作用,激发其荣誉感和责任心,使其乐于参加基金项目的评议工作。有些专家因学术水平高、知识面宽,对基金项目的评议认真负责,各相关项目主任都向其指派项目,致使这些专家收到的各类基金项目过多,有时高达50余项。专家评议基金项目过多时,其认真程度难免会有所下降。在一般情况下,项目主任给专家的评议时间大约是两周左右,而在此其间,专家还有正常工作和其他事情要做。如果专家对某项目很熟悉,那么看完申请书,写好评议意见,一个项目至少也需一、两个小时;如果还需要查阅文献,可能就需要三、四个小时或更多的时间。实践表明,一个专家评议十几项研究内容相近或相关的项目是可以接受的,但不同科学处的项目主任在信息互不沟通的情况下向同一评议人发送不同内容的项目且项目总数较多时,就很难保证评议的质量了。因此,如感到某专家的评议工作不够认真,项目主任还应做些调查,不能一概否认。随着自然科学基金会评议工作信息化进程的发展,项目主任将能够方便地获得评议专家所评项目数的情况,如果同行评议专家库足够完备,上述情况是完全可以避免的。

3 注意专家在评议过程中的公正性

同一领域的专家和申请人可能有千丝万缕的关系,如亲戚、朋友、师生、学派等。回避原则适用于至亲,但对于朋友、师生或学派等关系就不好掌握了。如果项目主任掌握的情况多一些,就容易处理一些,至少可以做到心中有数。统计和实际观察表明,大多数评议专家都比较公正,只有极少数评议人因与申请人关系好或有矛盾而故意拔高或贬低其申请项目。从学术上讲,小同行可以更准确地评价基金项

目,但有时小同行之间会有严重的利害冲突关系,具体表现在有些小同行专家对同类项目要么一味贬低、要么无保留支持,这就是同行评议中的“利益冲突”(Conflict of Interest)问题,是同行评议研究中不可回避的焦点之一,应引起我们重点关注。以下4点措施可能会在一定程度上有助于这一问题的解决。(1)提高资助率,可以适度降低“利益冲突”的程度,缓解其对评议人的影响;(2)增加每个项目同行评议人的数量,以减弱个别受“利益冲突”影响的意见对项目评议综合结果的影响;(3)坚持严格完善回避制度,对“同业竞争”程度激烈的评议人的意见要格外认真分析,可以消除一些较明显的影响;(4)在选择同行评议专家时适当选一些大同行,以有助于更客观地评价基金项目。另外,认真地对评议专家进行反评估也会帮助项目主任看出问题,从而指导以后的工作。当然,反评估的目的并不是要将评议人分成三六九等,而是更充分地了解评议人,以便今后的基金评审工作更加公正合理。

4 重视同行评议专家评议意见内容的有效性

评议专家返回的评议意见形形色色,参差不齐:有些评议专家没有具体意见,却有鲜明的评价和态度;有些评议专家有具体意见,却评价模糊、态度暧昧;有些评议专家既有具体意见又态度鲜明;有些评议意见洋洋洒洒却都是“套话”,少有针对性;有些评议意见寥寥数语却一针见血,明确点出项目的优势和缺陷。不论评议意见的形式如何,项目主任对那些能够直接具体地阐明项目的创新特色、以及存在的问题,从而对下一个评审环节有重要参考价值的评议意见都应给予特别的重视,这样的评议意见反馈给申请人,还可以对申请人在今后项目的执行或再申请方面起到重要的参考作用。如何才能使每一位评议人返回的意见都能如此有价值,是值得每一位项目主任认真思考和努力的。这首先要要求所选的评议人懂行,其次要求评议人评议的项目不能多。要保证这两点的关键因素是要有一个足够大而全面的专家库,并配以完善的管理系统。美国NSF每年受理5万多份申请,其同行评议专家库中有25万名专家,一方面保证有较宽的选择范围,同时限定每位专家评议的项目数,从其每年公布的情况看,评议7项以上的专家就非常少了。而自然科学基金会目前每年受理申请已达3.4万份,评议系统中的专家人数只有4.7万人,相比之下,明显不足。因此,必须

要给项目主任充分的条件使其能够尽可能多的了解各方面的专家,还应充分发挥评议系统的功能,允许各科研单位向评议系统专家库中推荐同行评议专家。但是,随着专家库人数的急速增多,按自然科学基金会目前的人力,仅靠项目主任的印象必然无法满足挑选评议专家的要求,尚需引入一些评估评议专家评议工作绩效的定量指标,以作辅助。

5 慎重科学地对待评议人评价项目的准确性

评议的目的就是选出好的项目,评议的准确性应是考察评议人的重要指标。但我们必须清醒地认识到,前面几点对评议人的反评估是定性的,而准确性则是定量的。在定量的过程中由于评价标准的不科学性会引入许多偏差,因此不能将它绝对化。采用什么样的参数来表示准确性,是反评估面临的重要课题。以命中率来衡量准确性,则项目最终是否获准资助就是标准,但这只是所有评审过程作用后的结果,并不一定就是客观的最好的结果。以成功率表示准确性即以资助项目是否达到预期目标为标准,需要多年的检验,且是否达到预期目标的标准难于把握,可操作性差,不便于实际中应用。若以离散率来表示准确性即以某评议人与其他评议人对某项目评价的偏差为标准,则这种标准与该项目的最终评价无关,恐怕也不够科学,且计算繁琐,工作量较大。经过认真研究和比较分析,我们在实际操作中选择了命中率作为主要指标来表示准确性。

某评议专家的命中率是指:其同意资助且最终批准资助与其不同意资助且最终未批准资助的项目数之和与其所评项目总数的百分比。它刻画了该评议专家的评议结果与最终评审结果的一致性程度。设 P_i 为评议专家 i 的命中率, m_{yi} 为评议专家 i 的评议结果与最终评审结果相一致的项目数, m_i 为评议专家 i 所评项目总数,由定义可知:

$$P_i = \left(\frac{m_{yi}}{m_i} \right) \times 100\% \quad (1)$$

通过近几年的实践,我们感觉采用命中率作为定量指标有其局限性,这主要是资助率带来的系统性影响未加适当考虑。目前基金项目的资助率在 20% 左右,有时更低。如果评议人对所评项目全盘否定则在平均情况下其命中率可达 80%,因此,如果评议人否定的多,其命中率就会偏高,如果赞成的多,其命中率就会偏低。在实际评审中,同行评议之后,还有项目主任综合和评审会评审两道环节。如果资

助率为 20%,上评审会项目约 30%—40%,则我们希望同行评议专家最好也筛选出 30% 到 40% 左右的项目。这样大多数评议人的命中率将有可能低于持全盘否定态度的评议人。另外,我们希望对创新性、跨学科交叉的项目给予鼓励,但如果某评议人对创新和交叉的宽容度较高的话,受最终资助率的影响,也有可能降低其命中率。因此有必要在命中率的表达式(1)中引入资助率 η 进行修正。但经过诸多尝试,我们始终没有找到一个合理的公式来引入 η 的修正,只能寄希望于资助率的逐步提高来缓解这一问题,另外也有必要引入其他定量指标对其进行补充。

6 推进基金评审信息化进程,逐步完善反评估指标体系

随着基金项目评审信息化进程的发展,评审工作的效率必将大大提高,同时也为反评估工作提供了更好的条件。现在我们已可引入更多的定量指标来对评议人进行尽可能全面的综合评价,以弥补命中率单一、定量指标的不足和片面性。我们设评议专家 i 对项目 j 的评议等级为 E_{ij} ,参与项目 j 评议的专家共有 n 位, n 位专家对项目 j 的平均评议等级为 \bar{E}_j ,则

$$\bar{E}_j = \frac{\sum_{i=1}^n E_{ij}}{n} \quad (2)$$

如果专家 i 共评议了 m_i 个项目,则我们有

$$\Delta_i = \frac{\sum_{j=1}^{m_i} (E_{ij} - \bar{E}_j)^2}{m_i} \quad (3)$$

我们还可以有

$$\delta_i = \sum_{j=1}^{m_i} (E_{ij} - \bar{E}_j)^2 \quad (4)$$

我们称 Δ_i 为专家 i 评议项目的标准方差,它表示专家 i 对各项目的评议等级与各项目的平均评议等级的偏差; δ_i 为专家 i 评议项目的算术偏差,理论上讲它会随着专家评议项目数的增多而在 0 附近波动。若(i 在正(或负)的方向逐渐增大或维持在一个较大的数值,则说明该专家在评价尺度的把握上相对偏松(或严)。

在今后的工作中,我们可以尝试将评议人的命中率 P_i 、标准方差 Δ_i 、算术偏差 δ_i 和总评议项目数 m_i 作为综合反映评议人评议工作准确性和经验的定量指标,以备指派评议专家时参考。我们必须注

意,准确性的评估是一种统计行为,评议的项目数越多,这些指标的参考价值越大,评议项目数如果较少,偶然涨落的影响就会增大。我们在第二点体会中已经讲到,专家的时间和精力有限,如果评议项目数过多,就会使其难于完成或敷衍了事,这样准确率也会降低。目前,我们一般将每位评议专家同期评议的项目数控制在5—15项之间。上述指标是可以逐年累积计算的,经过几年的积累,这些指标将会为我们提供可较客观评估评议人评议工作的参考数据,进而使评议专家库日趋完善。

表1列出了2002年半导体学科6位评议人的4项定量指标。因所评议项目数较少,因此只具说明意义。

表1 2002年评议中几位评议人的4项定量指标情况

评议人	评议项目数	算术偏差	标准方差	命中率(%)
1	12	4.9	0.70	58
2	12	-0.6	0.143	17
3	12	-7.6	0.73	83
4	11	-2.7	0.51	45
5	10	-3.8	0.82	50
6	10	-0.4	0.288	60

从表1中可以看出:第一位评议人标准方差为0.70,明显偏大,算术偏差为4.9,说明该评议人在评价等级的把握上有些偏松,但命中率尚可;第二位评议人虽然命中率只有17%,但其算术偏差和标准方差都非常小,说明该评议人在评价尺度的掌握上比较准确,在资助率约为20%的情况下,其态度总体上偏向于支持,实际上该评议人对其所评的12个项目均表示了支持,但如果将其评为特优和优的项目作为支持项目,而将其评为良的项目作为不支持项目,则该评议人的命中率可达92%;第三位评议人的命中率较高,但标准方差为0.73,明显偏大,算术偏差为-7.6,说明该评议人在评价等级的尺度把握上过于严格,另一方面由于资助率很低的原因,其命中率却相当高;第四位评议人标准方差为0.51,算术偏差为-2.7,显示该评议人评价尺度略显偏严,命中率为45%,有些偏低;第五位评议人标准方差为0.82,过大,算术偏差为-3.8,与第三和第四位评议人比较,说明该评议人可能有对个别项目的

评价过于偏低的情况,需要进一步分析调查,实际结果是该评议人对所评10个项目中的3个项目的评价低于其平均等级1至2个等级;第六位评议人的算术偏差和标准方差较低,命中率尚可,说明该评议人各方面尺度把握较好。这些例子说明采用上述4个指标可以使更好地了解评议人评议的综合特点,使我们在挑选评议人以及综合评议人的评议意见时心中更加有数。

总之,对同行评议专家的反评估是一项复杂而繁琐的工作。项目主任在综合评议意见时,对评议人的感觉是最深刻准确的,因此,定性评估的重要性和不可替代性是不言而喻的。事实上,每位项目主任对自己常年聘请的评议人都是有自己的认识的。定量评估的结果虽不及定性评估结果准确丰富,但也有其科学性和客观性,而且对项目主任有提示作用,也有助于这些关于评议人的宝贵资料在不同项目主任之间交流和传递。定量评估指标的选择非常复杂,加权难度大,例如,同样的命中率,对突出创新、增进交叉的非共识意见应该加权,在资助率很低的情况下还应考虑资助率的影响,不同的指标是否可以加权合并,等等,非常复杂,需要我们认真努力而且不断地探索。同时,定量评估是个复杂的系统工程,需要各方面的探索、协调和配合,如基金项目评审程序、评议表格和指标的设计、评议系统信息化的发展等。科学合理的评审程序、项目评议表格和指标,有利于简化对同行评议专家反评估的定量评估指标,并提高其科学性;完善的信息系统,则可以容纳较繁琐的定量评估体系。

以上几点体会是在前人工作的基础上获得的,有待于通过与同行的切磋和交流而进一步提高。最后,感谢董国轩、白坤朝、李晓民和吴卿等同志对拙作的讨论。

参 考 文 献

- [1] 吴述尧主编. 同行评议方法论. 北京:科学出版社,1996.
- [2] 商玉生主编. 十岁的思考. 北京: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] 夏传铨. 同行评议专家的经验权重. 中国科学基金管理专刊,1995.

SOME IDEAS ON EVALUATING THE REVIEW ACTIVITIES OF PEER-REVIEWERS

He Jie Wang Chenghong Liu Ke

(Department of Information Sciences, NSFC, Beijing 100085)