

· 科学论坛 ·

学科年度同行评议：基于遗传学与生物信息 学科基金项目评审意见的分析

谷瑞升^{1*} 张 暉² 王 胤³ 田艳艳¹ 杜生明¹

(1. 国家自然科学基金委员会生命科学部, 北京 100085;
2. 山东师范大学, 济南 250014; 3. 青岛大学, 青岛 266021)

[摘要] 同行评议是国家自然科学基金项目评审体系中最为重要的一环,也是决定资助项目质量的关键环节。了解同行评议状况,发现问题和改进不足,也成为完善专家系统和提升评审质量的有效途径。为此,在前期探索的基础上,我们从评议专家对项目的综合评价、对项目的熟悉程度、对项目评价的区分度、评议意见内容和评议专家间共识度等方面,对2016年遗传学与生物信息学科445位同行评议专家的评审意见进行了分析。结果显示,同行评议质量总体良好,绝大多数专家对评议项目把握准确,项目综合得分情况基本呈正态分布,97.8%的专家对所评议的项目熟悉/较熟悉,97.5%的专家对所评项目给出合理的区分度,98%的专家评议认真,撰写的评议意见具体并有实质性意见和建议。但也发现了极个别专家存在一定问题,如有的专家对评议项目缺乏区分度,有的专家学术判断存在较大问题,有的专家撰写的评议意见过于简短且俱为套话和空话。针对上述问题,我们与相关专家进行了沟通并给予督促,同时对有些专家在以后的同行评议中予以回避。本文对专家评议状况进行了分析,期于在管理上作出探索和尝试,为提高同行评议质量提供理论依据。

[关键词] 自然科学基金;同行评议;专家;区分度;共识度

同行评议是国家自然科学基金(简称科学基金)项目评审的第一步,也是评审体系中最重要的一环,是遴选项目的重要依据。同行评议优劣直接影响和决定了资助项目的质量,是国家自然科学基金科学与公正的基础。国家自然科学基金委员会(简称基金委)历来十分重视同行评议工作,经过30年的发展,建立了较为完善的同行评议体系,如相近研究采用捆绑式评议、送审最适宜评议的小同行等,并且不断探索和改进,努力提高同行评议质量,如提高专家库质量、建立申请项目和评审专家相一致的申请代码、研究方向和关键词系统、探索会评前网上投票遴选等。“十三五”期间,基金委更加强调与时俱进,卓越管理。杨卫主任在“贯彻发展理念,突出战略导向,不断提升创新驱动发展源头供给能力”的报告中强调了科学基金工作“深入落实五大发展理念,聚力

前瞻部署、聚力科学突破、聚力精准管理,完善资助体系,为实施创新驱动发展战略作出新贡献”^[1]。同行评议维护了科学基金的学术声誉,赢得了广泛赞誉。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,如极少数专家受社会上讲关系送人情等不良风气干扰,评议偏离了客观与公正;由于知识面和学识水平的局限,对项目的把握和判断出现偏差;项目评判过严或过宽;由于工作繁忙导致评议不认真或找他人替代等等。上述种种情况干扰和影响同行评议的公正性,急需着力去解决。同行评议是一个复杂的系统工程,专家的评议是一个综合判断的结果,既有对研究工作的评价,也有对研究者自身的判断,既受自身学识和研究动态把握的影响,也受周围非学术因素影响。因此,如何判定同行专家评议是否恰当显得非常困难。前人在这一领

收稿日期:2017-07-05;修回日期:2017-08-29

* 通信作者, Email: rsgu@nsfc.gov.cn

域也进行了探讨,如探讨同行专家的反评估方法^[2-4]、对改进科研项目同行评议工作提出了很好的意见和建议等^[5-7]。以上方法提出了很好的同行评议评估思路、方法和模型,但不足之处是多数仅针对同行评议中的一个方面进行评估。如何较全面和综合地了解同行评议状况,特别是甄别出有问题的评议和专家,优化评议专家队伍,是摆在进一步提高同行评议质量面前的一项十分重要工作。

本文依据科学基金项目内容相近指派给同一组同行专家的捆绑式评审方式,通过提取专家的评议信息,从对项目的熟悉程度、评议项目的区分度、评议专家间共识度、以及评议意见内容等几个方面,对专家的评议进行遴选和深入分析,从多角度了解专家的评议状况,为进一步采取有针对性措施,提高同行评议质量提供依据。

1 评估方法与评估内容

2016年遗传学与生物信息学科面上和青年项目申请共计907项,根据内容相近指派给一组同行专家的捆绑式评审方式,每个项目由5位专家进行评议,共有445位同行评议专家进行了项目评议,回收4535份评议意见(回收率100%)。在基金委ISIS评审系统中,通过输入评议专家姓名,获取同行专家的评议信息,根据评价指标进行整理。本文针对445位同行评议专家的4535份意见进行了深入分析,包括:项目综合得分分析、专家对项目熟悉度分析、专家评议项目的区分度分析、评议意见撰写与内容分析和专家共识度分析。

1.1 同行评议的总体状况

同行专家对项目综合判断体现在综合评价指标上,分为优(A)、良(B)、中(C)、差(D)。而根据现行的项目综合得分规则每位专家的每项得分的权重分别是:A为4,B为3,C为2,D为1。如果一个项目5位专家综合评价均为A,该项目综合得分为4,以此类推。根据综合得分了解本年度学科评议项目的总体情况。

(1) 专家对评议项目的熟悉程度。专家对项目的熟悉程度分熟悉、比较熟悉和不太熟悉,根据专家填写的信息对445位专家的4535份意见进行分析,了解专家对评议项目的熟悉程度。

(2) 专家评议项目的区分度。学科建议评议专家对所评议项目按照一定比例提出优先资助、同意资助和不建议资助建议,一般优先资助和同意资助比例控制在35%左右。本文针对445位专家的评

议信息进行了分析。

(3) 评议意见撰写和内容分析。对4535份评议意见进行了分析,遴选出较为简短的评议意见,依据意见查询到评议专家,认真分析专家对所有项目的评议情况,甄别出意见过于简短(少于120字)且套话空话无实质意见的专家,或评议意见在不同项目间高度相似的专家。

1.2 专家评议的共识度分析

共识度分析是基于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,符合度越高专家得分越低,反之专家得分越高。本文按照谷瑞升等的同行评议意见共识度分析方法^[7],对专家评议的共识度进行了分析。其基本思路是:提取专家对每个项目给予的综合评价指标,即给予每个项目的A、B、C、D。将定性综合评议指标数量化,通过与另外4位专家的综合评议指标比较,将量化指标标准化,利用数学方法对标准化数据进行分析,确定专家综合评价指标 S_i ,通过对综合评价指标进行排序,筛选出得分高和低的两类专家,并进行分析和验证。首先依据对专家的排序,选出排名前和后各10%—20%的专家进行单因子方差分析(One-Way ANOVA)统计分析,有统计学差异的共识度差的专家作为重点分析对象。然后从以下几方面进行分析。

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

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

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

通过以上分析找出重点分析和评估的专家,通过深入分析找出问题产生的原因。

2 结果与分析

2.1 同行评议的总体状况

2016年遗传学与生物信息学科面上和青年项目申请共计907项,综合得分情况见图1。从图中可以看出,项目综合得分基本呈现正态分布特点,从宏观上讲,对项目的判断是科学和合理的。按照约26.7%的资助率,242个项目应该获得资助,综合得分超过3.2的共计273项,综合评价均为良好的项目综合得分3.0,共计366项,资助项目需要优中选优。

(1) 专家对评议项目的熟悉程度。

对445位专家的4535份评议中熟悉程度统计发现,有2708份评议为熟悉,占59.7%,1724份评议为比较熟悉,占38.0%,另有102份评议为不太熟悉,约占2.3%,如图2所示。对不太熟悉的评议深入分析,发现受评审项目多为交叉性项目,如基因组、生物信息,多基因功能网络分析等。

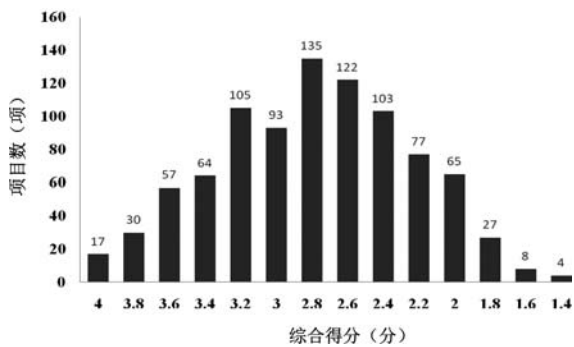


图1 2016年遗传学与生物信息学项目综合得分分布情况

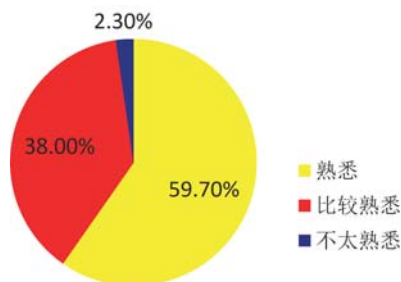


图2 2016年遗传学与生物信息学评审专家熟悉程度

(2) 专家评议项目的区分度。

445位同行评议专家中有434位专家对其评审项目给予了很好的区分,占比达97.5%。然而,有10位专家对所评议项目的综合评价均给予了A或B评价,即均同意资助,而有1位专家对所评议项目的综合评价均给出了C或D,即均不同意资助。这11位专家未对评议的项目给出合理的区分度,约占总评议专家的2.5%。

(3) 评议意见撰写和内容分析。

经过对4535份评议意见认真阅读,发现绝大多数专家(98%)具有科学认真的评审态度,对项目给予了客观评价,并提出了实质性的意见和建议,也遴选出了一些撰写过于简单空洞的评议。按照评议意见查询到评议专家,认真分析专家对所评议的所有项目评议意见情况,发现9位专家的评议意见较为简单且内容空洞,甚至有1位专家在不同项目上意见重复。

2.2 专家评议的共识度

为了更有效地提高评审质量,需要在更多方面对专家进行更细致的评估,包括评价尺度、学术判断力和客观公正性。本文采用专家综合评价指标 S_i 对2016年参加遗传学与生物信息学科项目评审的445位专家进行了评估。我们从总体上分析了445位专家对应的 S_i 及其分布,希望可以在总体水平上直观地观察他们的评审均一性与差异性,以方便从中选取 S_i 排名靠前和靠后的专家,以利于进一步的分析 and 验证。

总体上看,445位专家的 S_i 基本呈正态分布,主要分布在0.5—1.1区间,在该区间内的专家人数为416人,占总数的93.48%。在其余的29位专家中, S_i 小于0.5的为18人,占总数的4.04%; S_i 大于1.1的为11人,占总人数的2.47%,具体数据如图3所示。

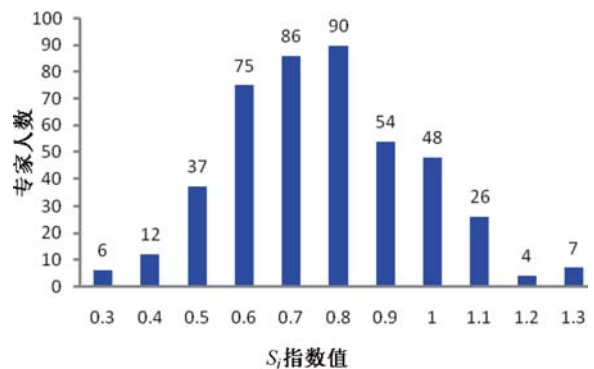


图3 2016年度遗传学与生物信息学科同行评议专家评估 S_i 分布结果

我们运用统计软件(SPSS 18.0)分别对排名靠前和靠后的30位专家的 S_i 进行单因子方差分析(One-way ANOVA分析),分析每位专家的 S_i 与所有专家平均 S_i 之间的差异,结果显示 S_i 靠前的30位专家中有15位具有统计学意义($P < 0.05$), S_i 靠后的30位专家中有13位具有统计学意义($P < 0.05$),针对具有显著性差异的专家,从同意资助率、优先资助项目的受资助率等几个方面做进一步分析

(表1)。

分析 S_i 值较低15位专家的数据,发现如下特点:(1)建议不资助项目的受资助率较低,在 S_i 值较低15位专家中有14位专家的建议不资助项目的受资助率为0,而 S_i 值较高的15位专家中有5位建议不资助项目的受资助率为20%—25%,明显偏高;(2)优先资助项目的受助率较高,在 S_i 值较低的15位专家中给予优先资助的11位专家中有6位专家的

表1 2016年度遗传学与生物信息学学科同行评议专家反评估结果

| 专家编号 | S_i (mean) | SEM | N | P value | 同意 资助率% | 优先资助 项目受 资助率% | 同意资助 项目受 资助率% | 不同意 资助项目 受资助率% |
|--------------|-----------------|--------|----|---------|------------|---------------------|---------------------|----------------------|
| S_i 前 15 位 | | | | | | | | |
| F1 | 1.3909 | 0.1398 | 8 | 0.002 | 12.50 | 100.00 | 100.00 | 0.00 |
| F2 | 1.3862 | 0.2163 | 7 | 0.004 | 28.57 | 50.00 | 50.00 | 20.00 |
| F3 | 1.3336 | 0.1585 | 9 | 0.003 | 55.56 | 50.00 | 60.00 | 0.00 |
| F4 | 1.3263 | 0.2365 | 10 | 0.002 | 50.00 | 33.33 | 20.00 | 0.00 |
| F5 | 1.3238 | 0.1504 | 7 | 0.011 | 28.57 | 50.00 | 50.00 | 20.00 |
| F6 | 1.3236 | 0.1342 | 8 | 0.007 | 100.00 | 66.67 | 25.00 | 0.00 |
| F7 | 1.3216 | 0.1820 | 8 | 0.007 | 37.50 | — | 33.33 | 20.00 |
| F8 | 1.2445 | 0.1675 | 9 | 0.015 | 88.89 | 40.00 | 25.00 | 0.00 |
| F9 | 1.2281 | 0.1880 | 7 | 0.039 | 85.71 | 66.67 | 33.33 | 0.00 |
| F10 | 1.1987 | 0.1061 | 15 | 0.005 | 46.67 | 50.00 | 28.57 | 0.00 |
| F11 | 1.1977 | 0.2390 | 9 | 0.030 | 44.44 | 66.67 | 50.00 | 0.00 |
| F12 | 1.1645 | 0.1592 | 11 | 0.028 | 27.27 | 50.00 | 33.33 | 25.00 |
| F13 | 1.1603 | 0.1603 | 10 | 0.039 | 60.00 | 50.00 | 33.33 | 0.00 |
| F14 | 1.1552 | 0.1760 | 10 | 0.042 | 50.00 | — | 60.00 | 0.00 |
| F15 | 1.1469 | 0.1214 | 14 | 0.019 | 64.29 | 50.00 | 44.44 | 20.00 |
| S_i 后 15 位 | | | | | | | | |
| L1 | 0.1279 | 0.0863 | 6 | 0.001 | 83.33 | 100.00 | 60.00 | 0.00 |
| L2 | 0.3069 | 0.0927 | 5 | 0.030 | 60.00 | — | 66.67 | 0.00 |
| L3 | 0.3142 | 0.1807 | 6 | 0.019 | 83.33 | 50.00 | 60.00 | 0.00 |
| L4 | 0.3772 | 0.0941 | 8 | 0.018 | 37.50 | 100.00 | 66.67 | 0.00 |
| L5 | 0.3861 | 0.1320 | 6 | 0.045 | 50.00 | 100.00 | 66.67 | 0.00 |
| L6 | 0.3941 | 0.1426 | 8 | 0.023 | 25.00 | — | 50.00 | 0.00 |
| L7 | 0.4106 | 0.1199 | 14 | 0.004 | 28.57 | — | 75.00 | 0.00 |
| L8 | 0.4333 | 0.2142 | 8 | 0.040 | 37.50 | 50.00 | 33.33 | 0.00 |
| L9 | 0.4451 | 0.1706 | 8 | 0.046 | 50.00 | — | 25.00 | 0.00 |
| L10 | 0.4504 | 0.1328 | 14 | 0.009 | 28.57 | 100.00 | 50.00 | 10.00 |
| L11 | 0.4527 | 0.1170 | 12 | 0.017 | 50.00 | 66.67 | 50.00 | 0.00 |
| L12 | 0.4605 | 0.1185 | 14 | 0.012 | 50.00 | 80.00 | 71.43 | 0.00 |
| L13 | 0.5016 | 0.1084 | 11 | 0.048 | 45.45 | 50.00 | 40.00 | 0.00 |
| L14 | 0.4774 | 0.1255 | 8 | 0.069 | 50.00 | 100.00 | 50.00 | 0.00 |
| L15 | 0.5262 | 0.1246 | 11 | 0.068 | 36.36 | 100.00 | 75.00 | 0.00 |

优先资助项目的受助率为100%,平均优先资助项目的受助率为81.52%,而在在 S_i 值前15位专家中选择优先资助的14位专家中仅有1位专家的优先资助项目的受助率为100%,平均优先资助项目的受助率为55.64%,经独立样本 t 检验分析,显示两者平均优先资助项目的受助率有统计学意义($P=0.04$),这说明 S_i 值较低的15位专家对于项目质量有着更清晰的区分和严谨的判断。

认真分析 S_i 值较高的15位专家的评议意见,发现以下几方面问题:(1)有些专家对项目要求过高,项目评价整体偏严格,没有很好地比较不同项目间的优劣和评价尺度,只是简单地否定如F1、F5和F12号专家的同意资助率偏低。(2)项目评价偏宽松,没有很好地比较不同项目间的优劣和评价尺度,只是简单地肯定,如F6、F8、F9和F13号专家的同意资助率均大于60%。(3)项目评审共识度较差,如F2、F5、F7、F12和F15号专家,一些建议不资助的项目最终获得了资助,这些专家学术判断力或公正性可能存在一定问题,与项目整体评价存在较大偏差。

3 讨论

本文从项目综合评价、专家对项目熟悉程度、意见撰写、区分度和共识度等方面较全面地对2016年参加遗传学与生物信息学学科项目同行评议的445位专家进行了评估,以共识度分析作为对专家评估的重中之重,通过对评议指标排序,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,较为客观、科学、综合和有效地对专家进行了评估,筛选出30位专家进行重点评估。结果发现15位专家的评议存在较为突出问题,同时也确认了13位共识度良好的评议专家,评估基本上反映出了专家整体评议状况,具有较好的科学性和可行性。

分析发现有10位评议专家的评议意见简单,无实质性内容,但该10人 S_i 并不处于前30位,说明这些专家虽然意见简单,但仍具有较好的学术判断力。10位评议人对所评项目均建议“优先资助/资助”,1位评议人对所评项目综合意见全部“中/差”,且全部为不资助,这10位专家对所评项目没有很好的区分度。该10位评议人中仅有1位 S_i 位于前30位,且有统计学意义;有1位 S_i 位于后30位,这可

能与评议专家所评议项目整体质量较好或较差以及评议项目数量有关。

笔者还对项目评议人的性别、职称(正高级与副高级)、年龄段(<35 、 $35-45$ 、 $46-55$ 、 >56)和工作单位类型(985与211高校、中科院系统、地方高校与科研院所)是否影响 S_i 分别进行了统计分析(数据未列出),结果发现上述因素对 S_i 影响均无显著性差异。

单纯地通过 S_i 值的排序也表现出一定的不足之处,因此我们进一步对 S_i 值进行统计分析,将统计学上的显著性差异分析与 S_i 值的高低结合在一起综合评价,根据差异显著性和 S_i 值的高低进行分级排序,以尽量减少评审专家在个别项目上产生的误差。尽管如此, S_i 值并不能保证评估结果完全正确。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二审专家、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评议人评审的情况;(3)项目资助与否,受所在领域影响较大,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;(5)共识度只是同行评议趋同性的一个指标,变革性项目或许成为同行评议趋同性中的另类,共识性评价指标的应用有很大的局限性。综上所述,根据 S_i 值并不能完全正确判断专家的评议状况,仅可以作为初选指标,还需要深入分析专家共识度差产生的原因,进而遴选出确因自身学术水平和学术判断或不公正等因素引起的共识度差的同行评议专家。

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。也希望为其他部门改进同行评议工作提供参考。我们也曾提议,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,不断地提高同行评议质量。

综上所述,本文对2016年遗传学与生物信息学

项目的同行评议状况进行了分析,总体情况较好;项目得分情况基本呈正态分布,有 97.8%专家对所评议的项目熟悉/较熟悉,97.5%的专家对所评项目给出合理的区分度,98%的专家对项目认真评议并给出建议,筛选出 15 位(3.4%)与其他专家共识度差且学术判断存在较大问题的专家。对学术判断存在问题和不公正的专家,我们在基金委评审系统中进行了标记,在以后的评审中将予以回避。

参 考 文 献

- [1] 杨卫. 贯彻发展理念突出战略导向不断提升创新驱动发展源头供给能力. 中国科学基金, 2016, 30(3): 193—196.
- [2] 谷瑞升, 张飞萍, 李永慈. 国家自然科学基金专家评议状况评估初探. 中国科学基金, 2005, 19(5): 298—301.
- [3] 何杰, 王成红, 刘克. 对同行评议专家评议工作进行评估的一些思考. 中国科学基金, 2004, 18(1): 47—50.
- [4] 赵黎明, 徐孝涵, 张卫东. 对同行评议专家的反评估分析. 中国科学基金, 1995, 9(1): 62—66.
- [5] 朱作言. 同行评议与科学自主性. 中国科学基金, 2004, 18(5): 257—260.
- [6] 曾畅. 完善科学基金项目同行评议体系的探讨. 研究与发展管理, 2007, 19(2): 124—128.
- [7] 江虎军, 冯雪莲, 杨新泉, 唐隆华, 何建庆. 影响科学基金项目同行评议质量的因素及改进措施. 中国科学基金, 2006, 20(6): 359—363.

Annual peer review status in the disciplinary of genetics and bioinformatics of NSFC

Gu Ruisheng¹ Zhang Liang² Wang Yin³ Tian Yanyan¹ Du Shengming¹

(1. Department of Life Sciences, NSFC, Beijing 100085; 2. Shandong Normal University, Jinan 250014;
3. Qingdao University, Qingdao 266021)

Abstract Peer review is the most essential in the project evaluation system in the National Natural Science Foundation, for it is a key step to influence and even decide the quality of the supported projects. In order to improve the quality of peer review, we should give an insight into the peer review, identify problems and improve deficiencies. Therefore, based on the preliminary exploration, we did an analysis on the evaluation results of 445 peer correspondents in the disciplinary of the genetics and bioinformatics in 2016 in terms of comprehensive evaluation, familiarity, discrimination, comments and consensus degree. The results showed that the quality of peer review was generally good, and most experts reviewed the project accurately and fairly. The degree of comprehensive evaluation presented a normal distribution. 97.8% of the experts were familiar with the contents of proposals, and 97.5% discriminated the proposals reasonably, and 98% gave specific and substantive comments. But we also found that very few experts had certain problems, such as lack of discrimination, judging differently with others, providing short and non-substantive comments. In view of above problems, we have taken some measures, such as communicating with the peer correspondents and giving them warnings, and avoiding those experts in the peer review system in the future. This paper provides a kind of exploration to analyze on the annual peer review, and might provide a base for improving the system.

Key words Natural Science Fund; peer review; expert; discrimination; consensus degree