

# 自然科学基金项目评价指标体系的研究

苏子仪\* 杨佐仪\* 陈国清\*

**【摘要】** 本文提出了建立评估指标应遵循的一些基本原则,确立了七大指标,并将其分为四个等级,均赋予相应的权数和数值,以便对条件不相上下的申报项目能更好地进行比较、区分,达到择优资助的目的。文章还提出了评审中应以定量为主,定性定量相结合的思想,既要便于面上大量的评审工作,且兼顾到少数具有突出个性的特好和特差项目,以提高评审结论的可信度,使评审工作朝着更加科学化、规范化和合理化的方向发展。

自然科学基金项目评价指标的确定是正常开展评审工作以及公平合理地选择资助最有价值的研究项目的重要前提,也是各申报者在同等条件下平等竞争的必要条件。因此,许多国家都根据国情认真研究和制定评审基金项目的准则。如美国就提出过四项选择准则:完成研究的能力;研究的固有价值;研究的有用性或相关性;研究对科学与工程学基础结构的影响。我国在同行评议意见书中也提出了四个方面,为定性地评估申报项目规定了范围和要求。有的科学部在评审过程中,为了对条件不相上下的项目能更好地进行比较、区分,以择优资助,又制定了8条或12条定量评估指标。从执行情况看,各有特点,但也存在某些共同性问题。本文试图在国内外已有的评价准则基础上,探讨建立评估指标的原则,确立评估指标以及定量评估的表格,以实现基金项目评审的规范化,逐步建立基金项目评审的计算机管理系统,其中包括申请项目的同行评议意见的数据库,具有运用表格形式的直观表达功能的软件以及评审专家库的检索系统。

## 一、设计评估指标的原则

评估指标的设计应遵循以下原则:

### 1. 非并合性原则

属于不同类型的指标不能相互并合,在制定指标时,不同方面的内容不能纳入一个指标之中,不应把不该包容在一起的内容包容在一起。如申请项目的目标、研究方法、技术路线等就需用不同指标来衡量。这如同不同量纲的物理量不能用一种尺度去度量一样。只有当它们之间具有并集关系时,才允许用一个指标代之。

### 2. 独立性原则

各指标内容不能相互隶属,相互推演。如技术路线的可行性分析和实现预期结果的可能性就不能成为彼此独立的指标。因为可行性分析就是对要实现的目标(即预期结果)所设计的技术路线以及所提出的步骤进行可行性分析,所以“实现预期结果的可能性”是与“可行性”重复的。从内容上看前者隶属于后者,因而前者作为一项指标不具有独立性。

\* 华中理工大学

### 3. 非并列性原则

在评审过程中,要给各项指标赋予不同的权重,主要指标和伴随指标不能并列,有主次之分。申请项目能否获得资助,重要的是取决于主要指标的水平。如“经费预算是否合理”,它就不是决定项目能否获得资助的决定性因素,而是项目获准后才伴随而来的问题,如果发现经费预算不够合理,可以建议修改。至于大型项目所申请的经费超过了基金资助的能力与限度,则可在项目指南中提出一个资助强度的波动区间。申请经费是否合理的问题,可作为参考指标,它可在项目排序中发挥作用。若两个申报项目评分一样,就看两者的经费预算谁更为合理,合理者就可排在另一个的前面。

### 4. 单义性原则

指标和指标的表述,概念都不能多义,必须满足单义性原则。

(1) 指标的内涵的理解问题。指标及其表述要明确,不能造成不同的评审者产生不同的理解,从而造成评审上的不一致,引起主观理解等因素参入作用,影响评审的客观性。如过去有人把“依据”理解为选题的客观根据(客观需要)与主观条件(包括实验设备,人员水平、能力,以及开题研究的起点等)。

(2) 对指标重要性的理解问题。若对指标的重要性理解不同,在整个基金委内评审标准就不一致。如不同的学科组赋予“选题的科学依据”的权数就很不一样。在生物学部的指标体系中,它占总分(60分)的 $\frac{1}{12}$ ,在材料学部则占总分(100分)的 $\frac{1}{20}$ 。这说明该项指标在不同学部理解是不同的。为了使评审工作规范化,提高评审水平,就应有统一的指标体系和量化尺度及权数。

### 5. 严格区分申请者的选题逻辑和评审者的评审逻辑

经分析,我们认为申请者的选题逻辑和评审者的评审逻辑是不同的,申请者是从客观需要(或问题)开始到实现目标的可能性的逻辑来选题的。当然,在这里选题的立论依据是首要的。然而,评审逻辑则是从预期的目标的意义、水平的评价开始的,因为如果这些方面较差,技术路线和方法再好也价值不大,不值得资助。所以在评审中,项目的预期结果和意义是首要的,相应的权重也高一些。因此,选题逻辑与评审逻辑是逆向的。显然,在评审中,应按评审逻辑来排序,而不能按选题逻辑来排序。指标的序位不同,权重也相应不同。

## 二、指标体系的确立

根据上述原则,我们建立起基金项目的评审指标体系。

### 1. 指标体系的内在逻辑

我们根据对以往评审指标的研究和专家意见的聚合,提出了七大指标,并将其分成四个等级(见表1),且均赋予权数和数值,不过它不显示在表上,而是在计算机程序中。第一等级为一、二项。这两项是首要的,因为如果选题意义不大,目标的水平不高,即使技术路线、方法再好,也不值得资助。所以这两项赋值最高(30分),指标的序位也排在前列。第二步考虑实现目标的可能性和条件,研究其技术路线及方法是否能保证项目的实现。这两项指标属于第二等级赋值(25分)。第三步考察执行技术路线的人和物等条件是否具备,研究者及合作者的水平与能力如何,这项指标属于第三等级赋值(20分)。如果第一、二等级的指标都得到满足的申

请项目, 仅人员配备得尚不够合理, 评审者可建议进行调整或补充。第四等级赋值(15分)的指标是研究基础和实验条件。实验条件不是特别重要的, 当然具备此条件更好, 不过我们不是要求每个单位的实验条件都要搞得小而全。如果实验条件不具备, 可通过协作或逐步加以补充解决。这四个等级的指标就是涉及到项目本身的主客观条件及实现它的可能。

表1 同行评议意见表

评议内容	评议标准	在同意的等级上划√
一、研究内容的科学价值或应用前景	1. 处于科学前沿, 有重大科学意义或有重大的应用前景	A <sub>+</sub> A A <sub>-</sub>
	2. 有较大的科学意义或有较大的应用前景	B <sub>+</sub> B B <sub>-</sub>
	3. 有一定的科学意义或应用前景一般	C <sub>+</sub> C C <sub>-</sub>
	4. 科学意义不大或应用前景较小	D <sub>+</sub> D D <sub>-</sub>
二、研究项目的目标评价	5. 属于该领域的国际先进水平的研究	A <sub>+</sub> A A <sub>-</sub>
	6. 处于该领域的国内领先或国际一般水平	B <sub>+</sub> B B <sub>-</sub>
	7. 处于该领域的国内较先进水平	C <sub>+</sub> C C <sub>-</sub>
	8. 目标不明确或水平一般	D <sub>+</sub> D D <sub>-</sub>
三、研究方法的先进性	9. 方法上有重大突破	A <sub>+</sub> A A <sub>-</sub>
	10. 现有方法移植, 有一定创新	B <sub>+</sub> B B <sub>-</sub>
	11. 方法一般	C <sub>+</sub> C C <sub>-</sub>
四、技术路线的可行性分析	12. 采用方法不恰当	D <sub>+</sub> D D <sub>-</sub>
	13. 技术路线清楚, 有独创性, 有一定风险	A <sub>+</sub> A A <sub>-</sub>
	14. 技术路线清楚, 方案明确, 具体、可行	B <sub>+</sub> B B <sub>-</sub>
	15. 技术路线、方案不够具体	C <sub>+</sub> C C <sub>-</sub>
五、申请者和合作者的研究能力和水平	16. 技术路线不太清楚, 方案不明确, 难于实现	D <sub>+</sub> D D <sub>-</sub>
	17. 有较高学术和技术水平, 能胜任该项目研究	A <sub>+</sub> A A <sub>-</sub>
	18. 有一定的学术和技术水平, 基本上能胜任该项研究	B <sub>+</sub> B B <sub>-</sub>
	19. 申请者有一定学术和技术水平, 但人员结构不太合理	C <sub>+</sub> C C <sub>-</sub>
六、研究基础	20. 水平一般, 完成课题有困难	D <sub>+</sub> D D <sub>-</sub>
	21. 该项目已开展研究, 已取得初步研究成果	A <sub>+</sub> A A <sub>-</sub>
	22. 已开展初步研究, 取得一定的资料积累	B <sub>+</sub> B B <sub>-</sub>
	23. 研究刚刚开始, 资料积累不太多	C <sub>+</sub> C C <sub>-</sub>
七、实验条件	24. 研究尚未起步, 无研究工作积累	D <sub>+</sub> D D <sub>-</sub>
	25. 具备完成该项目实验条件	A <sub>+</sub> A A <sub>-</sub>
	26. 实验条件基本具备	B <sub>+</sub> B B <sub>-</sub>
	27. 实验条件不太具备	C <sub>+</sub> C C <sub>-</sub>
计算机累计的结果	21. 无完成项目的实验条件	D <sub>+</sub> D D <sub>-</sub>

## 2. 关于指标量化问题

这里有两个问题需要进一步研究, 一是量化的必要性, 二是量化的可能性。评估指标量化, 既可克服由于自然语言本身所具有的某种模糊性, 导致人们理解上的一些不可避免的偏差, 又可对评议后的项目进行比较, 能大致按照优劣排序, 尽可能使评审工作做到科学化、规范化、合理化。并可避免评审人的不必要的重复劳动, 促使评审者对每项评审内容都予以考虑, 并作出自己的评价。从而可避免因考虑不全面或其它某种原因, 而在书面定性评审意见中, 出

现忽略一些必要的评审内容的情况。纯定性评估,虽然使用丰富的自然语言,能突出最优和最劣的项目,但对大量情况(条件)类似的项目,用自然语言书写的评审结论,常常具有模糊性。如对“好”的程度理解往往就不完全一致,而且这类定性评审结论具有不可比性,不可测性。从现在执行的情况看,这类项目的定性评审结论,语言已趋于模式化,进而可进行形式化和量化。这说明对项目进行定量化评估已有了可能和基础。所以我们的同行评议意见书是采用以定量为基础(即大量项目的评审结论应以定量评估为依据),定性与定量相结合(即突出好的项目和特别差的项目,在定量评估的基础上,重点参照定性评估栏的结论)的方式。

### 3. 表格的制定

(1) 制定表格的要求:(a)科学。指标制定得要具有科学性,能反映项目的水平;(b)冗余度小。要求反映项目水平的指标做到无冗余度,避免出现无用的和重复的指标;(c)简明。要求各指标的定性描述简单明了,不让评议专家产生歧意理解;(d)使用方便。专家评审时,只须在等级栏内对同意的打√,不计算分值。而且尽可能采用以整数赋值,易于计算。

(2) 表1中将“研究内容的科学价值”和“应用前景”放在同一项指标栏内,即为“研究内容的科学价值或应用前景”。这表示,评审人已判定被评议的项目属于“基础研究”,那么该项指标就指“科学价值”。若判定被评议的项目属于“应用研究中的基础性工作”,则该表的这一栏指标主要是指应用前景。科学价值与应用前景两者在这里是等价的。

(3) 每一子指标均分为三个等级,如:A<sup>+</sup>, A, A<sup>-</sup>; B<sup>+</sup>, B, B<sup>-</sup>,等级分得比较细。从以往评审情况看,特好和特差的项目的占少数,中等档次的项目占多数。如果等级太少、太粗,就易使项目得分很集中,给确定资助对象带来困难。指标等级设计得较细,可使各申报项目得分距离拉开,便于对项目排序,便于确定资助起点和资助项目。

(4) 关于申报项目是否重复的问题,我们没有把它确定为决定项目能否获准的判决性指标之一。原因是:(a)项目的选题是否重复,从其它指标可体现出来。如某项目的预期目标(或成果)的指标,达到或接近世界先进水平,或为国内先进水平,这就表征了该课题不具有重复性。(b)确实为重复课题,可在形式审查中淘汰。申报课题与公布于众的研究成果一般是不会重复的,谁都不愿做重复的劳动,而且其成果也不会被社会所承认。但是申报课题有可能与已开题而尚未完成的课题重复。在目前尚未建立全国性课题查重档案的条件下,自然科学基金资助的课题是有条件查重的,一旦属于纯重复性课题,在形式审查中就应淘汰,不必留到同行评议阶段进行评审。(c)如果形式审查中不能处理,进入同行专家评议阶段就应慎重审查。从不同角度重复某课题,应该是允许的,这就要以最后取得之成果来竞争优劣。又若课题性质或目标有所重复,但有新的技术思想或方法,这仍然是可取的。若在课题申报过程中,有两个同时申报的课题重复,这时就应象对其他课题评审一样,从各项指标比较优劣,决定取舍。可见,“重复性”在同行评议阶段,并不是决定课题能否获准的关键性因素。因此,没有将它确立为指标。

(5) 指标的赋值。我们根据元老法,请有较深造诣的专家对各项指标的重要程度给出分值(实际为重要程度排序的人数),我们把这称为暂定重要系数。然后据此再求出各项指标的重要系数(即权重),最后再由给出的总分值(160分)和各项指标的重要系数,求出各项指标应得总分(见表2)。

表 2

指标编号	暂定重要系数	重要系数%	分值
1	60	18.75	$160 \times 18.75\% = 30$
2	60	18.75	30
3	50	15.625	25
4	50	15.625	25
5	40	12.5	20
6	30	9.375	15
7	30	9.375	15
合计	320	100	160

取总分为 160 分的原因在于:(a)使各大指标之间的比例不会悬殊太大;(b)使各大指标根据权重数可取整数赋值,并使计算简便;(c)为满足各申请项目排序的需要。总分高,等级可分得细一些,以拉开各项目得分的距离,便于项目排序。再根据资助总金额和支持强度,就可确定参与排序的项目获准的起分线。(d)满足心理因素的需要。这样的赋值可使一般项目所得总分不会太低,对申请者给予心理上的安慰和鼓励,增强再次申请的信心。

### 三、计算机应用于基金管理方面的两点设想

1. 我们就表格生成和专家定量评估意见聚合等问题,如何采用计算机来实现,作了一些初步的探索。在表 1 中,只要求同行评议专家对各项指标在同意的等级栏内打“√”,而不必打分或进行分数累加,专家主要从定性方面对项目作出综合评价,量化工作由计算机按程序进行计分、累加、排序。为了便于基金委在管理上实行宏观控制和专家查询,我们还打算研究计算机自动表格生成,按照基金项目评审工作的要求,显示申请者姓名、单位、项目编号、项目名称及所得总分。并且对获得相同总分的项目,可用计算机汉字显示评审表的情况,以便比较,决定取舍。此程序进一步与专家系统连网,可构成一个大系统。

2. 我们设计的同行评议表中,对申请项目,既有定性综合评估,又有定量指标的各项评估和最后累计。这就便于对定性评估与定量评估进行对比分析。若两者偏差不大,就说明可用定量评估逐步取代定性评估的某些功能,使评审工作走向规范化,实现计算机参与管理工作。定性与定量相结合,以定量评估为基础的方法,将提高评审结果的可信度。

## STUDY ON THE TARGET SYSTEM FOR EVALUATING NSFC PROJECTS

Su Ziyi, Yang Zuoyi and Chen Guoqing

(Huazhong University of Science and Technology)

### Abstract

The article presents some basic principles for establishing the evaluation targets and enumerates seven major targets which are further divided into four grades and given corre-

sponding weighting coefficients and values. Thus, the submitted projects which are in more or less the same levels may be better compared and distinguished and their best selected ones will get the support. The article also puts forward the idea of stressing the quantitative method while combining it with the qualitative method in the evaluation work. Consideration will not only be given to many general projects but also the small number of best and worst projects with special characters. In this way the evaluation will develop in a more scientific, standardized and rational way.