

# 关于如何评估数学研究人员的建议

2014年8月11日国际数学家联盟全体大会讨论通过

## 一. 引言

如何评估和对待学者的研究工作？这是从大学出现之初到如今一直反复出现的问题。这个问题与科学家的社会价值和教育的社会功用密切相关，因此我们应当在更广泛的背景下去讨论这个问题。许多评估流程并不限于数学和数学家：每门学科面临着类似问题。因此，跨学科的课题组多年来已就这些问题提出许多文件/指南。因此本文刻意不着重讨论永恒的和重要的问题，这些问题已被充分彻底地探讨过（例如关于利益回避规则等），也不考虑对所有学科都适用的问题。本文目标是探讨以下两个具体问题：

\*哪些方面的评价是数学所特有的？聚焦这个问题的原因是数学常常需要与其他学科区别对待，对许多学科而言团队合作和基金资助具有决定性作用。数学家们经常发现自己难以在跨学科圆桌讨论中说服其他科学家，讨论结果往往令数学家感到失落。本文旨在为数学家在与其他学科的同事辩论中提供帮助。

\*最近的社会进步（互联网，国际化，科学团体的扩张，经费制约，审计规则泛化）对评价数学家的副作用是什么？哪些负面影响需要纠正或冲抵？

当我们强调数学的特殊性，不应忘记数学界有相当大的重要组成部分，由于明显和正当的原因，其工作方式与纯粹从事学术研究的数学家略有不同。对于这些数学家，我们在本文下一段落将要讨论的“标准”而具体的评价办法（主要基于对公开发表论文进行的详细考察）还须重新调整。例如，对于涉及保密条款、工业应用、或者软件开发项目的应用数学家，对于跨学科研究或从事数学教育的数学家，我们需要一种不同的视角。因此，本文分为两部分：第一部分是关于学术研究型数学家，第二部分讨论其他类型数学家的评价问题。

**重要声明：**学者业绩的评价可以有多种不同目的（聘用，晋升，经费等），并且在大多数情况，研究活动的评估只是众多考核指标之一。保持良好的学术环境需要学者多方面的能力和素养，在做决定时需要考察多方面的能力和素养。本文只讨论评估中关于研究活动的那部分；在附录中，我们给出一些事例，说明可用于评价个人学术水准的其他标准。

## 二. 学术研究型数学家

由于数学涵盖范围极大而各分支特色迥异，半自动化评估可能误判，评估标准不具有普遍性，统一的评价标准也未必是我们追求的目标。例如：

\*不同国家的大学系统是不同的，不同学术背景的各类数学家增进了国际学术团体的多样性。

\*遴选国际大奖获得者和选聘青年教师的评估方法是不一样的。

\*全国性学术社团可能考虑采取某些特别措施，例如，当他们察觉潜在的弱点并希望纠正（可以是，在某些领域缺乏创新和原创性或在其他领域缺乏严密性和清晰度），因此在做决定时要进行评估。

现如今评估委员会通常考察

\*个人著作目录（包括期刊的名称；合作者，论文页数）；

- \*个人陈述，用不太过于专业化的语言介绍自己的研究工作；
- \*一位或多位同行专家撰写的评议，这些同行应已阅读当事人的某些论文。

这些专家可以是评估委员会成员，也可以是外请的，他们的作用至关重要。事实上，为了正确评价论文的价值和意义，需要真正懂行的专家仔细检视这些论文。基于杂志影响因子的半自动化定量评估很容易导致错误。乍一看，这些方法看起来客观、科学、不被人为操纵、不受争议左右。然而，事实并非如此，它们可以带来一些非常消极的副作用：

\*高水平的研究需要原创性、发明发现、以及与之相伴而来风险（启动一项雄心勃勃的研究计划时没有任何成功的保证）。所有这些因素都很难在基于文献计量的标准化评估中得以体现。

\*文献的量化评价导致论文数的激增，因为这种评估偏好系列论文，其中的结果是逐步完善的。一般而言，对于数学研究，更重要的是精心撰写的、无可改进的、最后形式的论文。从长远来看，一篇富有新意的论文，无论发表在哪种期刊，对学术界的影响远超十篇技巧性的普通论文。事实上，预印本服务器如今可以发布尚未投稿的论文初稿，这些初稿经过一段时间的改进，成为一篇更详细或更简洁或更明确的论文后才投稿发表。

\*影响因子：一篇论文已被引用多少次？查看这样的数据也可以很有意思，数学社团本身就提供这样的数据，例如美国数学会的 MathSciNet，但引用这些数据应慎之又慎。首先，数学的某些领域与别的领域相比往往发表更多的论文，我们因此不能把某个领域工作（如范畴论）的个人与另一个领域（如生物统计学）工作的个人比较这些数据。第二，人为推高文献计量数据（例如通过交叉引用等）或操纵影响因子并非难事。

因此，尤其需要鼓励在评估委员会任职的数学家向评估委员会的其他学科成员说明基于文献计量学的评价对数学家特别不适用。我们希望本文可以帮助做好这一点。同时我们必须强调，数学家并不主张其他学科应该改变他们的具体评价标准；国际数学联盟并不知道评估化学家或经济学家的最佳方式。本段的结论是下面明显的表述，它是本文的核心：

**没有任何方法（尤其是涉及数字和数据的半自动化的伪科学的评估）可以替代由真正了解的专家撰写的同行评议。此外，就数学研究而言影响因子之类的指标工具显然是无用或无关的。**

一个可能看起来很诱人的方案是引入更科学的文献计量学的替代标准，须知目前最具影响力的文献检索工具都是商业公司编撰的，（对他们来说，这是一项商业活动），但给定这样的文献工具内在的负面影响，国际数学联盟并不希望参与这样的工作。

### **审计理念与科学实践，评价活动的爆炸性增长。**

近几十年来“评价”的作用变得越来越重要。发端于商业环境中（会计及经营）的“审计”概念，现已渗透到西方社会的方方面面。人们相信可以对几乎所有的事务、个人、组织、企业、产品等进行统一的、具有可比性的、客观的评价。特别是许多资助机构现在完全相信评估科学活动的重要性和普适性，他们往往会坚持使用自己制定的评价规则，往往是以半自动化的“客观”标准，如关键绩效指标(KPI)。而多数专家认为这些方法一般都不能很好地适用于评估科学活动，特别是数学。

这种“泛化的审计理念”的另一负面影响是评价活动过多过泛，因为每一决策层都希望进行自己主导的评价。很显然，某些评价活动是必要的，并且原则上每位活跃的数学家都能花费一些时间用于评价其他人的工作，为杂志审稿，正是最重要的和最必需的部分。同时，他们应保留尽可能多的时间进行他们自己的研究。评价活动的泛滥是真正的危害。它还诱导科学家本身定位的变化：如何进行科研，如何报告成果。研究的目的是为了推进我们对科学规律的认识，而不是为了获得良好评价。偏离这一目标将导致非常消极的后果。

### **较小的科学团体。**

如何使所有科学团体都能用上一个健全、合理的评价程序？找到适当的评估专家？较小的国家，或者那些研究活动非常庞杂多样的科学团体，在评议研究成果的档次和质量时，很难获得可靠和客观的信息。似乎人们经常面临两难选择：只依靠当地学术团体（明显有自我评价和利益冲突的危险—应当明确被禁止），或半自动化的文献计量系统。如上所述，长远来看这并不是解决问题的好方法。人们需要另一种解决方案。

是否可以建立一个超越国家的组织来帮助开展评价活动？国际数学联盟有无责任为落实这一想法而发挥作用？赞成的理由包括：还有谁能做此事？反对的理由是，国际数学联盟的主要目标应该是促进数学家团结合作，而不是制造紧张，而评估活动不可避免地产生矛盾。国际数学联盟直接介入可能得不偿失。

## **三. 其他具体意见**

### **多学科和工业数学。**

如上所述，评价数学家应基于对其科研成果的认真评估，不宜采用任何一种半自动的刻板计量(如 KPI)。在本段，我们把目光投向参与跨学科研究的数学家，无论是在学术界还是在产业界，对他们评价具有特殊性。

这里只限于讨论这样的数学家，他们接受其他学科和工业（或其他行业）应用的挑战，发展了新的数学理论，并用来解决实际问题。建模—研究—开发—应用，这一工作流程与纯学术的数学研究有点不同。我们再次强调，研究工作如果只涉及现有数学工具或技术的直接应用，不在本段的讨论范围（这种类型的工作，可根据相关应用领域中的标准进行评估）。

由于跨学科研究和工业界的发表论文的风气差异挺大，丰富多样，对这类研究活动的评价更需依赖同行专家评审，非常苛刻而又非做不可。鉴于此类评估活动的重要性，无论是数学本身还是应用方面，评估需要精心组织实施。

除了前面关于数学创新的评价准则以外，我们还应进一步采用其他标准来评判一些额外的和具体的挑战：

\*从问题来源方的角度来考察数学对解决实际问题的益处和作用是什么？这一点非常重要。我们因此不必十分在意数学细节的完整性或理论意义的重要性（不要与缺乏严谨性和新颖性相混淆），这些缺失可由相关合作领域的其他重要指标来补偿。例如，在其他学科，论文的第一作者具有很强的含义，而按字母顺序署名是数学论文的传统。总结一下：这样活动的评估可包括其他学科中使用的标准，但数学本身的评判标准也是必不可少的。

\*在业界或工业项目中数学家还面临其他问题。某些特别限制条款导致研究工作不能全文发表，或受知识产权限制，（更经常地）或因缺乏时间给出全面详细的证明。评议组或评议者应当考虑这一现实情况。

\*与此相关的一个问题是，这类研究和开发活动的“最终产品”不一定是论文，它可以是软件，而编制和实施这样一个蕴含数学性质的软件可以是非常基础性的和耗费时间的。这个例子说明，进行评估时应考虑到重要贡献的各种可能形式。

## 数学教育

下面一些简短评论是关于如何评价从事数学教育的研究人员。从一个国家到另一个国家，这类学术团体的组织结构差异很大，其成员有时来自大学或学术机构，有时隶属于教师协会，有时候则是教育部的一部分。此外，它们的目标、定位、方法手段和活动内容可以相差很大，从认识论角度分析如何讲授数学知识(通常在学术范围内)，短期课堂实验的设计与分析(通常涉及教师)，设计和分析师资培训课程，或者依托学校进行大型的调查研究，不一而足。对不同类别的学生(普通学生或天资优异)分类培养，或者不同的社会和文化背景，也都是需要注意的因素，并且需要认知科学和社会科学的投入。为了提高所有层次数学教学和学习质量这一最终目标，所有这些学科都是需要的。因此，本领域涉及一些数学，更需要前述的各类学科知识。因此评估有关成就需要融合各个领域的标准。

### 附录：评估中应该考察的方面,一份不完全清单：

学术活动的所有方面都是不可或缺的，学术团体需要这些方方面面的活动才能正常运转。它们也可以间接地，但以显著方式，推进高水平研究。请注意这是一个相当长的清单，我们并不指望有人勾选所有选项（譬如，编写计算机软只涉及数学界的一小部分，并非所有的研究人员参与教学工作等）。

- \*发表在国际期刊上研究论文;
- \*研究专著、教材、课堂笔记;
- \*软件和程序代码的编写及应用;
- \*组织学术活动，尤其是在学术机构通过竞争性评选获得批准的活动;
- \*组织学术会议和研讨会，尤其是在学术机构通过竞争性评估选定的提案;
- \*讲授课程，开设新课，教学成果奖;
- \*评审论文;
- \*学术奖励;
- \*指导学生：博士生、硕士生、本科生、进修教师;
- \*在学会或学术机构当选为资深会员或院士;
- \*咨询活动，包括国际学术期刊编辑委员会工作;
- \*宣传活动：科普文章，公众演讲，在学校开设讲座或竞赛，教师协会任职等;
- \*行政事务;
- \*效率和可靠性。

本文件是由国际数学联盟执行委员会设立的工作小组撰写的，工作小组由四位数学家组成：

- \* Mariolina Bartolini Bussi (数学教学国际委员会(ICMI)提名)
- \* Carlos Cabrelli (发展中国家委员会(CDC)提名)
- \* Andreas Schuppert (由工业与应用数学国际委员会(ICIAM)提名)
- \* Wendelin Werner (组长，代表 IMU 执行委员会)

章复熹译 陈大岳校