

学术评价的元评价机制

朱少强^a, 唐林^b, 柯青^a

(广东工业大学 a. 管理学院; b. 图书馆, 广东 广州 510520)

摘要:元评价是对评价的评价,它对于规范各类评价活动,发现评价中存在的问题和偏差,提高评价质量具有重要作用。当前中国学术评价体系中存在一定的混乱现象,需要建立相应的元评价机制来加以引导和规范。可以考虑从对评审专家的评价、对评价机构的评价、评价行业协会、学术出版认证等角度来构建中国学术评价体系的元评价机制。

关键词:元评价;学术评价;规范化;认证;行业协会

中图分类号:C931 **文献标志码:**A **文章编号:**1008-5831(2010)03-0081-05

评价的根本目的是服务于管理,是要为管理者提供辅助决策的关键信息。尤其是当今“大科学”时代,学术评价的结果几乎都会直接影响各学术机构和学术从业人员的切身利益及各种资源的分配。自20世纪90年代以来,随着以量化评价为代表的各种评价考核方式的大规模应用,由学术评价引起的争议也越来越大。人们逐渐认识到,学术评价能否发挥预期的作用,很大程度上取决于评价自身的质量。某种意义上,评价是一柄双刃剑,可能促进也可能扭曲学术事业的发展。因此,对于规范和衡量学术活动成效的学术评价自身,也很有规范和衡量其好坏的必要。这就涉及到所谓的“元评价”问题。

一、“元评价”概念及对学术评价的元评价

“元评价”即“对评价的评价”,其最初提起及应用主要是在教育领域。“元评价”概念最早是由美国评价理论专家麦克·斯克里温(Michael Scriven)在1969年的教育产品报告中首次引入的,当时是为了描述一个对教育产品设计的再评价之评价,即对评价进行再评价的过程。国外学者对于元评价比较权威的解释是:“元评价一般是指对评价技术的质量及其结论进行评价的各种活动。元评价就是对于评价的评价。其目的是向原来的评价者们提出他们工作中存在的问题和片面观点。”^[1]在中国,对评价自身质量问题的研究是从20世纪80年代中期开始的,中国学者把元评价又称为“元评估”、“再评估”、“再评价”、“后设评价”等^[2]。元评价的重要意义在于:对评价活动自身的好坏进行反思,总结成功经验,纠正评价过程中的失误和偏差,有利于实现评价活动的科学化和规范化,从而提高评价活动的质量或适用性。可以认为,元评价是评价活动从纯粹的实践探索,逐步走向理性化、科学化和系统化的一种表现形式。

收稿日期:2010-03-12

基金项目:国家社会科学基金项目“网络信息资源评价指标体系的构建与实证分析”(08CTQ015);广东工业大学博士基金项目“基于自组织模式的网络信息资源评价研究”(083045)

作者简介:朱少强(1975-),男,湖北黄冈人,广东工业大学管理学院讲师,博士,主要从事信息管理与科学评价研究;唐林(1975-),男,湖南永州人,广东工业大学图书馆馆员,主要从事信息资源管理研究;柯青(1979-),男,安徽黄山人,广东工业大学管理学院讲师,博士研究生,主要从事知识管理研究。

欢迎访问重庆大学期刊社 <http://qks.cqu.edu.cn>

如同对评价活动自身的关注一样,可操作性和规范性正在成为元评价的核心话题。1981年、1988年美国教育评价标准联合委员会先后公布了《教育方案、计划、材料评价的专业标准》和《人事评价标准》。1991年由美国教育统计局资助的合作性教育数据收集和报告标准项目工作小组公布了《教育数据收集与报告标准》。1994年由美国教育评价标准合作委员会制定了“项目评价标准”,1995年由美国评价协会制定了“美国评价协会指导原则”,这是判断项目评价质量的两个主要工具,它们之间相互协调和补充,共同作为项目评价元评价标准的基础。1999年,丹尼尔·斯塔弗尔比姆在西密歇根大学评价中心以教育评价标准合作委员会的项目评价标准为基础,开发了项目评价的元评价标准检核表,包括四个标准,即实效性标准(Utility)、可行性标准(Feasibility)、适合性标准(Propriety)和准确性标准(Accuracy)。元评价人员可以根据该检核表和量化的公式,采用定量计算和定性描述相结合的方法,综合得出科学、公正的元评价结论^[1]。

关于元评价的内容和方法,国内外诸多学者进行了有益的探讨。例如,王从双等认为,以具体评价活动为对象的元评价主要有以下内容和标准:一是对评价方案进行评价,包括评价对象和评价目的、评价标准、评价指标和权重分配、评价方法、信息收集和处理方法等几个方面;二是对评价的组织实施进行评价,包括对评审专家组成员和评价组织工作的再评价;三是对评价的结果进行评价,包括评价结果被接受的程度,评价的信度与效度等^[2]。王敏提出内容分析法、经验总结法、评价信度分析、评价效度分析等四种元评价方法^[3]。马宁锋等按评价活动中的各种要素来划分元评价的对象,将元评价分为对评价主体的评价、对评价内容的评价、对评价方法的评价、对评价结果的评价四种类型^[4]。钱存阳等用多元统计分析中的克伦巴赫(L. J. Cronbach) α 法和因子分析来分别检验评价体系的信度和结构效度^[5]。冯晖借鉴软件工程中软件测试的原理方法,提出了基于黑盒测试和白盒测试的元评价方法^[6]。杨毅等采用斯皮尔曼等级相关系数对科技进步综合评价指标进行分类,确定强相关和弱相关指标,借以分析在指标设立方面存在的问题,优化指标体系等^[7]。

目前国内关于“元评价”的研究文献主要集中在教育评价领域,学术评价领域内涉及“元评价”的极少。其主要原因可能在于:目前中国学术评价的理论和实践发展,还远远没有教育评价那样成熟,还处于实践摸索的初级阶段,而缺少科学性和规范性。

但近来随着对学术进行量化考核引起的争议越来越大,一些学者也开始认识到“元评价”之对于学术评价的重要意义。例如在“人文社会科学评价问题”学术研讨会上,与会专家认为“与评价问题相关的还有‘对评价的评价’问题。由于普遍存在的评价程序不公正,使评价常常出现‘外行评内行’现象,必须对评价者进行审定。要建立相互制约、彼此监督的社会机制来解决对评价的再评价问题”^[8]。当前学术评价实践中的“工具至上”思维十分流行,管理操作的观点主导了评价实践;对于评价主体、评价目的、评价标准认识不清,各种评价方法、指标任意套用、胡乱搬用,象把数量当水平(以发文数量评教授)、把投入当效益(科研项目 and 科研经费的多少),或者笼而统之搞“综合”的现象比较多;由于缺少科学和规范的理论指导,造成许多问题、争议。因此与教育评价领域相比,当前中国学术评价体系也迫切需要建立一种元评价的机制。

由于元评价是“对评价的评价”,即以评价活动自身为评价对象的。而根据对评价活动的构成要素进行分析,一项完整的学术评价活动,应该包括2个实体性要素——评价主体、评价对象,以及8个非实体性要素——评价目的、评价标准、评价内容、评价程序、评价方法、评价指标、评价数据、评价结果等的参与;其中评价主体又分为委托方、评价方、被评价方三方^[9]。因此,可以考虑以评价活动的10个要素为线索,构建学术评价的元评价体系和机制,例如对评价主体的元评价,对评价方法和指标的元评价,对评价标准和数据来源的元评价等。

二、对评审专家的元评价

评价方是贯穿整个评价活动的主要实施者,“元评价”的主要目的是对评价方建立约束机制。如果说程序公正机制是着眼于建立一种基于“过程”的约束机制;那么“元评价”则主要着眼基于“结果”的约束机制。评价主体的各方相对独立、分离,并处于一个委托/代理关系的链条中;这是实施“元评价”的基本逻辑前提。

最重要的元评价机制,可能是对同行评审专家的评价。权利与责任对等,这是实现公正、公平的基本原则;然而现有的同行专家评审体制,基本上是有权、无责。以“双盲”通讯评审为例,看起来似乎很公正,但这要建立在选对了真正的同行专家,且专家认真负责、人品高尚可信赖的假设基础之上。同行专家手里握着无形的权力,掌握着学术成果、研究人员乃至重点学科、重点研究基地等的命运;但除了良心以外,对专家本人似乎没有更多的约束——就算专家把通讯评审的工作交给他的学生来做,或者假公

济私、搞人情关系,或者潦草塞责、随便应付,既难以明眼看出,也难以采取惩罚措施,因为专家毕竟有权发表自己的任何意见。随机遴选专家的规则只是降低了搞人情关系现象的可能,并不能解决专家是否称职、是否有能力胜任、是否认真负责的问题。理论上,也可以公开评价过程与结果,如果评价结果不公,可以提出申诉,但也只是再另外找人复审而已,作弊的专家一般并不会有什么损失。因此,需要对评审专家建立一种事后的、基于信誉的约束机制。

对同行评审专家的“元评价”,卜卫等人提出了一系列方法,包括评估者整体信度分析、评委个人信度分析、评委严格度分析、评委个人一致性检验等。其基本思想是:信度分析的任务实质上是测量评委间态度的一致性。而每个评委的信度,即是该评委与其它评委态度的一致程度。评委间对成果评价的一致性程度越高,就认为这个评委组信度越高——而这可以用统计学的方法来加以判别。评估者整体信度用肯德尔系数 W 或卡方值 χ^2 估计。评委个人信度用相关系数 R 或协方差 COV 估计。评委严格度用评委打分均值与总体均值之差来代表。如果评委对各成果打分未能保持基本一致的宽严程度,则判定该评委在某些成果打分上可能不公正;该评委对该成果的打分应在汇总计时去掉^[10]。笔者认为,除上述分析之外,对评审专家的专业水平、评审态度、信度等信息,还应该纳入一个可以在学术界共享的、保存有连续累积数据的数据库系统中,纳入到一个信誉、信用管理机制当中。如果某评审专家有多次不良信用记录,则他的评审权利就将被取消。这样才能真正形成“元评价”的机制。

三、第三方评价机制和对评价机构的元评价

评价机构也是重要的评价主体。由于评价活动是由评价机构具体组织,有关评审的操作流程、方法和指标一般都是由具体执行评价的机构来设计,评审专家也多数是由该机构来聘请,所以评价机构的专业水平如何、组织是否健全等,对于评价的科学性和公正性有极大的影响。对评价机构进行元评价的主要障碍在于:当前作为“评价方”的评价机构和作为“委托方”的科研行政管理部门基本上是两位一体的,致使评价活动及其评价机构没有相对独立性,妨碍元评价机制的发挥。

对于评价机构的“元评价”,首先需要市场化的、多元竞争的环境。发展独立的“民间”第三方评价机构,委托方基于这些机构的专业水平、特长、信誉等,与其中某一机构签订委托评价合同。委托方可以对这些中立的第三方评价机构建立信用、信誉管理体系。只要有充分的市场竞争存在,市场本身就是以

成为评价机构的“元评价”机制,就会自动对评价机构构成强有力的约束。

如前所述,掌握科学研究资源分配的科研行政管理部 门,虽然在形式上是直接的评价委托方,但根据公共行政理念,它自身也是作为科学共同体或公众的“代理人”而执行其使命。因此,更有必要把“评价”的职能从科研行政管理部门中独立,以接受社会监督和保障公平。基于学术评价过程中行政干预因素过多,如搞“平衡”、讲“照顾”等情况;许多学者建议,应大力发展独立的第三方评价机构,许多评价活动可以交给这类机构去做,既把科研行政管理部门从具体的评价组织活动中解放出来,又可以保证评价的独立性、避免外来因素的干扰。按照市场经济逻辑,有垄断就有不公,有竞争才有约束。应该发展一批独立的第三方评价机构,以便相互形成竞争局面,而不是另设新的官僚机构和形成新的垄断。“官方”的评价机构和民间的评价机构可以并存,并在一个评价行业协会的认证和监管之下,做到相互监督、相互竞争和行业自律。

四、评价行业协会和评价活动的规范化

如果第三方评价机构比较繁荣、规模相对较大,还需要建立科学研究评价的行业协会,用以对科学研究的评价标准、评价办法、评价指标、评价主体资格、评价程序等实施认证、规范等活动,进行行业信息披露、行业自律和监管。评价行业协会可以代替所有委托方,对众多第三方评价机构及同行评审专家统一信用管理。与竞争性市场一样,行业协会也是一种重要的元评价机制。

对评价指标、评价方法、评价程序进行标准化、规范化,也是评价行业协会的一项重要“元评价”职能。目前国内各单位学术评价往往各行其是,没有统一的标准和操作规范;从许多关于“指标打分法”的研究来看,依靠个人主观感觉、好恶设计评价指标体系的很多。一些评价体系试用的结果,缺少严格的实验数据支撑,仅仅是感觉试用效果“良好”。这些指标体系、方法,没有经过长期、多次的试验,很可能并不成熟,就在实践中使用,难免会造成问题。而建立元评价机制,对这些评价程序、方法、指标形成稳定的指导性规范,将有助于解决当前学术评价中出现的种种问题。

当前国内学术界的功利与浮躁情绪,不仅表现在评价指标的数量化导向上,也反映在评价方法、指标本身的形成程序上。蒋国华曾经指出,早期的诸计量学量化指标主要是用于为了科学决策而开展的科学质量和科学发展趋势的研究,也就是宏观层面的应用;应用于科学家个人成就的评估,是后来的

事^[11]。在科学计量学的理论研究方面,西方发达国家自始至终走在前列;而在实践运用科研计量评价方法方面,前东欧社会主义国家却更加积极、激进,西方发达国家则显得保守。至今为止,美国、英国仍然实行以同行评议为主的科学研究评价体制。笔者认为,西方发达国家这种“慢决策,快执行”的谨慎态度,值得我们学习。

谨慎的态度,要求有相应的“元评价”机制。对于一切量化评价方法、指标,以及同行评议的方法、程序,或者如数据库、网络等现代化技术手段,不管新的或旧的,应该持一种开放的、批判的态度。在不同的环境下,允许不同的评价方法、指标进行一定程度的试验,但在得到公认、发展成熟之前,使用应该谨慎。在不同的评价机构、评价方法、评价指标、评价标准、评价程序之间,应该存在一种开放、有序的竞争机制,最终让优秀者在竞争中脱颖而出。这种开放的环境和态度,本身就是一种“元评价”机制。但与此同时,需要警惕“工具至上”思维和简单的“拿来主义”,认真区分研究性评价和实践性评价;注意实践性评价的要求,对于任何应用于大规模评价实践的评价指标体系、评价方法、评价程序等,一定要经过实验的反复检验,证明比较成熟、可靠才行。为此,建立类似评价行业协会等有形的“元评价”体系也是极有必要的,其作用就在于对于各类评价机构、评价专家、评价方法、评价程序等进行研究、管理、规范和认证,加强评价活动本身的规范性。

五、学术出版认证机制和对评价标准的元评价

学术规范的普遍遵循是定量指标评价方法有效的逻辑前提。否则,即便采用看似客观、公允的定量指标评价方法,也不能挽救学术评价的失效和学术研究的失范^[9]。大力加强学术规范,改善学术生态环境,不仅是科学研究事业健康繁荣发展的需要,同时也是完善学术评价体系的需要。但学术环境、学术规范的改善,是一项复杂的系统工程,非一朝一夕之功。当务之急是把好“杠杆”——学术成果发表规范,以求“四两拨千斤”的效果。具体地说,就是要加强学术论文、著作“双盲”评审的同行评议审稿规范;同时,仿照 ISO9000 等认证体系,对学术期刊、学术出版社进行学术认证。

学术期刊之所以是“学术”期刊,不仅仅因为它刊载科学研究论文;同时也因为它决定论文发表的过程符合学术规范,即所有发表的科学研究论文都通过了学术同行的严格审查,并认定其确有创新价值。这才是“学术”期刊的真正奥义。对学术期刊的论文稿件实行双向匿名评审,一直是国际上通行的做法,并已久经考验,运作良好。1997年,中国政法

大学张保生副校长作为中国高校社会科学代表团成员访问美国加州大学伯克利分校,一位期刊主编在介绍匿名审稿制时反复强调:匿名审稿制是发现高水平论文的唯一有效办法。没有匿名审稿制,就不可能有高水平的科学研究^[12]。但目前由于种种原因,匿名审稿制并没有在我国得到普遍遵守,许多学术期刊没有足够的经费支持,或本身办刊就是为了赚钱而不是为了学术;出于办刊成本等考虑,没有稳定的审稿专家,仅仅靠编辑把关,制度不严格、不健全,人情关系稿泛滥,发文随意性很大。这就严重损害了学术成果发表的应有程序规范,给学术评价也造成了困难。为此,应在学术期刊发文及出版社出版学术著作过程中,大力推行基于同行评议的双向匿名审稿制度,充分发挥同行评议作为学术“看门人”的角色。

为了增强各学术期刊、出版社贯彻匿名审稿制的积极性,作为辅助措施,可以仿照美国各大学院系进行专业认证的做法,建立基于品牌、信誉管理的学术期刊认证制度。如果某学术期刊拥有相对稳定、高水平的评审专家队伍,且在决定论文发表过程中严格遵循匿名审稿制度,就将之列入认证名单中,以表明该期刊具有可靠的学术水准。否则,就从认证名单中撤销。这种做法并不稀奇,如 SCI 源刊的说明中,就显著地标明该刊是否经过同行评议。

六、结语

元评价就是对评价体系或评价活动本身的再评价,它有利于从体制上规范各类评价体系和评价活动,发现评价实践中存在的问题和偏差,提高评价质量。元评价机制的建立,是一类评价活动走向成熟和规范的标志。当前我国各类学术评价活动比较丰富,但以量化考核为代表的评价体系引起了一定的混乱和争议,亟需建立相应的元评价来对学术评价活动加以引导和规范。从对评审专家的评价、从评价机构的评价、评价行业协会、学术出版认证等角度来建构学术评价的元评价机制,将有助于推动学术评价事业的健康发展,从长远来看也将进一步推动人文学术的繁荣。

参考文献:

- [1] 高洁,蔡敏. 美国教育评价的元评价及其启示[J]. 教育研究,2007(6):21-24.
- [2] 王从双,向德全,董彦省等. 教育评价的元评价初探[J]. 空军工程大学学报(军事科学版),2005(3):51-53,64.
- [3] 王敏. 对大学课程评价之“元评价”的思考[J]. 云梦学刊,2008(5):120-121.
- [4] 马宁锋,李卫东,薛艳泽. 基础教育学业成绩元评价的探

- 析[J]. 中国电力教育,2009(8):29-30.
- [5] 钱存阳,李丹青. 多元统计分析在课堂教学质量评价元中的应用[J]. 数理统计与管理,2005(6):40-43.
- [6] 冯晖. 基于测试的元评价技术[J]. 上海师范大学学报(哲学社会科学版),2007(5):128-132.
- [7] 杨毅,张晓东. 基于元评价的科技进步综合评价指标分类研究[J]. 哈尔滨商业大学学报(社会科学版),2007(3):78-81.
- [8] 刘劲杨,刘永谋. “人文社会科学评价问题”学术研讨会综述[J]. 中国人民大学学报,2004(2):155-156.
- [9] 朱少强,张洋. 科研评价的信息过程——基于“世界3”理论和“代理”假设的解释[J]. 科学学研究,2009(9):1305-1311.
- [10] 卜卫,周海宏,刘晓红. 社会科学成果价值评估[M]. 北京:社会科学文献出版社,1999.5:171-182.
- [11] 蒋国华. 科学学的起源[M]. 石家庄:河北教育出版社,2001:249-250.
- [12] 张保生. 论程序正义与学术评审制度的建构[J]. 学术界,2001(6):148-157.

Meta-evaluation Mechanism of Academic Evaluation

ZHU Shao-qiang^a, TANG Lin^b, KE Qing^a

(*a. Management School; b. Library, Guangdong University of Technology, Guangzhou 510520, P. R. China*)

Abstract: Meta-evaluation, namely the evaluation of evaluation, is favorable for the standardization and improvement of evaluation activity, correcting biases as well. There is a demand for meta-evaluation mechanism because some confusion exists in current academic evaluation system. The evaluation of evaluators, the evaluation of evaluating corporations, the guild of academic evaluation, and academic publisher authentication are preferred solutions to initiate the meta-evaluation mechanism for academic evaluation system.

Key words: meta-evaluation; academic evaluation; standardization; authentication; guild

(责任编辑 彭建国)