·本刊特稿·

“三大核心”:拿什么来取而代之?

——学术评价的困境

朱 剑

(《南京大学学报》编辑部,江苏 南京 210093)

[摘 要]学术界和学术期刊界与评价机构的“缠斗”已不下二十年,之所以难分胜负,是因为评价机构有着自己的理论支撑和数据支持,其以“三大核心”为代表的排行榜得到了行政权力事实上的采信;但因专业能力的缺乏,其理论、数据和算法及结果都有经不起推敲之处,其排行榜在学术评价实践中的运用更是导致了一系列严重问题。在现行科研体制下,行政权力既不能无视排行榜带来的问题,但在资源分配与科研管理中又无法舍弃它,故而左右为难,举棋不定,而代价就是学术评价陷于困境、学术研究遭受伤害。清理“四唯”“五唯”行动迈出了重建学术评价机制实质性的一步,但关键还在于能否找到“三大核心”的替代品,而被寄予希望的“定量评价与定性评价相结合”和“代表作评价制度”因条件所限很难取而代之。在这一背景下“三大核心”制作者的进路选择也值得关注。

[关键词]学术评价;核心期刊;CSSCI;定性评价;定量评价;代表作评价制度

[中图分类号]G236 [文献标识码]A [文章编号]1671-3842(2019)02-0005-21

一

在中国学术界和学术期刊界,学术评价是一个热议多年的问题了,而提起学术评价,必然会说到“三大核心”。所谓“三大核心”,一般指的是:北京大学图书馆等单位制作的《中文核心期刊要目总览》、中国社会科学院文献信息中心制作的《中国人文社会科学核心期刊要览》(今已被中国社会科学评价研究院《中国人文社会科学期刊评价报告》取代),以及南京大学社科评价中心制作的“中文社会科学引文索引(CSSCI)来源期刊目录”。说白了,就是从1990年代初到1990年代末先后问世的三个学术期刊排行榜,其发布者无一不是以文献情报学人员为主组成的评价机构。虽然仅仅是几页或几十页纸的榜单,其能量却不容小觑,因为在几乎所有由行政权力主导或组织的学术评价中,这些榜单都起到了不可替代的作用,故而说其牵动了整个学术界和学术期刊界并不为过。

围绕“三大核心”,有两个现象想来学术界和学术期刊界都已不再陌生。其一,自从“三大核心”问世不久,学术界特别是学术期刊界对它的批评和讨伐之声就没有平息过,特别是遇有新版期刊排行榜发布之时,讨伐之声就会形成一个令人瞩目的高潮。笔者将其称为学术评价的“潮汐现象”:低潮时常有,而每隔一段时间,必然来场震动学术界和管理层的“天文大潮”。其二,尽管对评价机构期刊排行榜的讨伐从来也没有中断过,讨伐的言辞也不可谓不激烈,然而却收效甚微,“三大核心”的期刊排行榜照样高调推出,以此为蓝本的各高校和科研机构的考核指标日趋刚性。于是,学者和期刊人对这些排行榜的迎合也就不可避免。迎合出现的标志就是对评价机构偏好的研

[基金项目]本文系教育部人文社会科学研究规划基金项目“高校学术期刊发展战略研究”(项目编号:14YJAZH116)系列成果之一。

[作者简介]朱剑,《南京大学学报》原执行主编。

究,从事这种“研究”的行列正在不断地扩展之中,从学术期刊界波及了整个学术界。

学术界和学术期刊界与评价机构的“缠斗”不下二十年了,似乎已达成一种巧妙的动态平衡。按说就这么几个评价机构,百十个核心制作人员,如何能够与数以百万计的学者和数以万计的学术期刊人相抗衡,并使得后者中的许多人对其刻意迎合?秘密就在于学术评价背后的行政权力。

在当今的科研体制下,行政权力不仅垄断了几乎所有学术资源的分配权,而且完全掌控了科学研究的管理权。在这场权力游戏中,行政权力充当的是组织者和裁判员的角色,而学者扮演的则是运动员的角色。权力的运作需要有科学的依据,亦即裁判员需要依据标准对运动员的成绩作出判定,这一过程实际上就是学术评价;继而就是组织者根据评价结果对运动员予以奖励或处罚,这一过程实际上就是资源分配和科研管理。资源总是相对稀缺的,行政权力要让有限的学术资源发挥出最大的作用,就要找到最合适的学者或科研单位来承接资源,此后还要对资源的使用进行管理。所以,只有先有评价结果,才能进行后续的处置(分配和管理)。可见,行政权力作用于学术研究时,至少包括评价权和处置权,前者是后者的基础和依据,而要行使好这两个权力,除了需受职业道德约束之外,裁判员是否具有专业水准和眼光之于评判结果是否科学和公正就显得十分关键。这也是在竞技界裁判员大多出身于运动员或专业研究人员的道理所在。但是,行政权力并非由专业学者所组成,亦即行政权力的行使者并不专业,这就意味着他们是无法直接进行评判的,如此,就只能将评价权与处置权分立,行政权力不得不将评价权分离出来委托给有能力进行评价的人或机构来行使。当然,评价权只是为处置权服务的,只要找到合适的受托人或机构就可以。

那么,什么样的人或机构有资格和能力接受行政权力的委托,承担起评价的责任,提供合适的评价结果呢?行政权力对评价结果的要求无非有二:一是权威性,即评价结果因其科学和公正而具有不可质疑性;二是简明性,即评价结果对评价对象孰优孰劣一目了然,从而使据此进行的处置或决策具有可操作性,最简明的当然莫过于排行榜。然而,要同时满足这两个要求是极其困难的。学术研究尤其是人文社会科学研究的复杂性决定了见仁见智乃是常态,随着各学科一言九鼎式的开创者逐渐退出研究领域,不可质疑的权威性的评价更显其难;同样,人文社会科学研究的复杂性也决定了其评价结果本身的复杂性,而不是排行榜的形式所能呈现的。

在1990年代之前,行政权力所需要的评价往往是委托给各种常设的或临时的专业委员会的,这些委员会大多由各学科的知名专家组成,所实行的评价方法多是同行评议,这也符合国际学术界的惯例。但是,随着国家对学术研究投入的加大以及学术生态的恶化,同行评议的公正性和公开性遭到的质疑日益增多,学术界对更为公正和客观的评价的呼声开始高涨。正是在这样的情况下,以客观和公正相标榜的量化评价出现在人们的视野,“三大核心”则是其代表。与同行评议相比,“三大核心”似乎更加符合行政权力对评价的要求,虽然其制作者说不上权威,但其以客观数据说话的话语形式也是一种对科学和公正的权威性的诠释,而以排行榜为产品形式正好满足了行政权力对评价结果简明性的需要。因此,到21世纪初,在由行政权力主持的学术资源分配和科研管理中,“三大核心”几乎成为学术评价的唯一主角,其制作者也已当仁不让地担负起了学术评价的责任。当然,在这一切的背后,是行政权力对同行评议的抛弃和对“三大核心”为代表的量化评价的采信。

“三大核心”对同行评议的取代并非一蹴而就,而是至少有一个从1990年代初到21世纪初这大约十多年的过程,随着这一进程,其弊端也开始显现:一方面,“三大核心”已逐渐成为学术界和学术期刊界的“指挥棒”,但这一“指挥棒”所引领的方向与学术研究和学术期刊的发展规律并不相符,严重干扰了正常的研究和办刊工作;另一方面,学术评价从来没有像现在这样与学术利益紧密勾连,而评价方法又是如此的粗暴和单一:只看排行榜,不上榜不仅意味着无法获取资源,甚至意味着出局。人们不难发现,从1990年代末以来不断增多的学术不端(剽窃、造假和粗制滥造等)事件背后,都可以找到量化评价的影子。于是,对以“三大核心”为代表的量化评价的批评和讨伐之声逐渐响起,并由零星的“发声”演变为学术界和学术期刊界的“大合唱”。

不断的批评和讨伐表明学术界和学术期刊界并不甘心听命和受制于这样的“指挥棒”,希望通过批评和讨伐来改善甚至摆脱这样的指挥。但是,“不管每一版期刊排行榜问世时批评和讨伐之声有多猛烈,三两年后,下一版排行榜仍会在更猛烈的批评和讨伐声中隆重推出并继续走红,而关于学术评价问题的热议还会无休止地继续下去。”①这是笔者七年前在一篇关于学术评价的小文中写下的一段话。撰写此文时,因 2010 – 2012 年版 CSSCI 来源期刊目录公布而引发的一场讨伐风暴刚刚平息。在笔者看来,这固然不是学术评价机构遭遇的第一场,也决不会是最后一场讨伐风暴。果然不幸而言中,相信学术界对于 2017 – 2018 年版 CSSCI 来源期刊目录公示后某些学术期刊主编发起的讨伐风暴一定记忆犹新。讨伐者的非理性出击和评价机构的不予正面回应与以往如出一辙;同样,这场声势不可谓小的风暴随着其他热点事件的出现也逐渐淡出了人们的视野,而更多的学者不得不加入到迎合的队伍中来。

学术界和学术期刊界与评价机构的“缠斗”之所以持续二十余年而不见止境,是因为作为资源分配者和管理者的行政权力态度的暧昧:既不干预学术界和学术期刊界对以“三大核心”为代表的评价机构的批评和讨伐,也不中止在学术资源分配和科研管理中对“三大核心”事实上的运用,从而造成了学术界和学术期刊界对评价机构的讨伐与迎合的奇妙组合:不讨伐,于心不甘;不迎合,利益受损。这样的状况的持续至少说明了以下几点:

第一,在行政权力主导的学术资源分配和科研管理中,学术评价有着无法替代的作用,行政权力通过对学术评价委托权的掌控深度地介入了学术研究,从而也使得评价由一般的学术权利而蜕变为名副其实的评价权力。

第二,伴随这一蜕变的是量化评价对同行评议的取代,继而是排行榜对量化评价的取代。前者剥夺了学术共同体从事学术评价的权利;后者让专门从事排行榜制作的评价机构垄断了学术评价。

第三,任何评价机构要获得评价权力,都必须使其产品得到行政权力部门的采信。尽管到目前为止,尚未有一家评价机构或其产品得到高层行政权力的公开承认,亦即尚未有任何评价机构正式获得行政权力关于学术评价的委托权,但作为学术评价产品的“三大核心”已在各级行政权力主导的学术资源分配和科研管理中发挥了决定性的作用却是不争的事实。这一事实说明,在既有的各种学术评价活动中,“三大核心”已与权威的学术评价画上了等号,其制作者因其产品在学术资源分配和科研管理中的决定作用已跻身于学术权力场中。

第四,在资源和利益均来自行政权力的当下,体制中人对权力的迎合是必然的,而当评价成为一种权力时,对权力的迎合首先就是对评价的迎合,对评价的迎合同样具有必然性。在现实中,对评价权力迎合的最典型的表现就是对“三大核心”偏好的研究并投其所好。

第五,学术界和学术期刊界对“三大核心”的刻意迎合行为说明了其背后往往是某种程度的屈从。如果评价完全符合学术研究和学术期刊的规律,就不会存在研究评价机构的偏好而削足适履的迎合行为。既然有屈从,就必然有反抗,所以迎合和讨伐才会同时呈现。

由以上几点观之,“三大核心”之所以得到行政权力的加持,是因为它们在一定程度上满足了行政权力分配和管理学术资源的需要;同时,来自学术界和学术期刊界前仆后继、锲而不舍的讨伐又意味着“三大核心”存在着重大缺陷。这就是行政权力部门既不抛弃“三大核心”,又不向其公开授权,也不阻止学术界和学术期刊界讨伐的原因。于是,“缠斗”也就达到了一种微妙的动态平衡。那么,这样的讨伐与迎合的平衡何时会被打破?笔者不敢断言,但当时间来到 2018 – 2019 年之时,几乎同时发生的两件事至少会对这一平衡产生一定的影响,或者还会掀起一番前所未有的波澜。

其一是评价机构新版排行榜的集中发布。如果不出意外,新一版即 2019 – 2020 年版 CSSCI 来

①朱剑:《重建学术评价机制的逻辑起点——从“核心期刊”“来源期刊”排行榜谈起》,《清华大学学报》(哲学社会科学版),2012 年第 1 期。

源期刊目录即将发布,而此前不久,另两大"核心"刚刚发布了新版排行榜。不论哪个"核心",新一版榜单的发布,有进有出是必然的,而评刊标准、方法也都或多或少会有变化,从已发布的两个"核心"来看,谁进谁出固然引人注目,但后者显然更值得关注。"三大核心"在相距很短的时间内分别发布新版已是多年不见的现象了,这种现象引发"天文大潮"当是大概率事件,但至少从目前来看,超大的风暴潮尚未如期而至,当然,最终是否会呈现,还要看新版"C 刊"发布后的情况。

其二是始于 2018 年中、当下正席卷高校和科研单位的清理"四唯"(唯论文、唯职称、唯学历、唯奖项)或"五唯"(唯论文、唯帽子、唯职称、唯学历、唯奖项)行动。所谓"唯",大概是只看一点,不及其余的意思,也就是只以某一项指标为评判标准,而不管其他,"三大核心"显然具有这样的特征。虽然此次行动并不专门针对"三大核心",但"三大核心"必然受到波及则是确定无疑的。

以上这两件同一时段中发生的事情颇值得玩味。此次清理"四唯""五唯"行动表明,行政权力一反过去对排行榜的暧昧态度,对科研体制特别是评价机制的改革似乎要动真格了。此项行动的开展,对陷于"缠斗"中的学者和学术期刊人来说,多年来的努力和付出总算没有白费,平衡即将打破,胜利的天平已向他们倾斜。而对评价机构来说,"三大核心"在这个时候发布新版榜单不大会是对清理"四唯""五唯"行动的公开挑战,必须看行政权力脸色行事的制作者还没这个胆量,只是按照惯例正巧到了该更新的时候,箭在弦上不得不发而已。但在这样的高压态势下发布新版榜单,要说没有压力是不可能的。所以,"三大核心"作出策略上的应变也是必然的。

那么,"三大核心"在学术评价中的唯一或核心地位是否会就此轰然倒塌呢?对此,笔者以为问题并不如此简单。清理"四唯""五唯"行动涉及的只是评价的标准和方法,并不触及由行政权力掌控资源分配和科研管理的科研体制,也就是说,学术评价在资源分配和科研管理中的地位不会因此而改变,能改变的只是委托谁来从事学术评价。清理"四唯""五唯"行动无疑是一种"破",所要"破"的固然是在学术评价中以类似"三大核心"这样的排行榜为唯一标准的做法,"破"字确已当头了,但是,"立",即对评价机制的重建,具体而言,找到"三大核心"的替代品以作为学术评价的新标准,是否也就在其中了呢?换言之,我们能否找到"三大核心"的替代品,就成为决定此次清理"四唯""五唯"行动能否取得预期效果的关键。本文将对此展开分析,不当之处,敬请方家指正。

二

要回答"三大核心"能否被取代的问题,我们先要弄清"三大核心"成为"核心"而号令天下这一局面是怎样形成的。一言以蔽之,是因为它们被行政权力相中而加持了。那么,为何这三个目录荣幸入选呢?一个重要原因就是,在参与学术评价之初,"三大核心"制作者们就为自己精心地打造了"科学"的学术评价者的形象,为其排行榜披上了"公平""公正"的外衣,颇具迷惑性。

首先,不管哪个"核心",在其背后,都有一套精心构造的评价理论支撑着。这套理论至少可以追溯到 1930 年代英国文献计量学家布拉德福(S. C. Bradford)发现的文献集中与分散定律。布氏发现,某一学科或主题的论文都相对集中地分布于少数期刊中,这些期刊就形成了对该学科或主题最有贡献的核心区。接着,在 1950 年代,美国文献计量学家尤金·加菲尔德(Eugene Garfield)发现了期刊论文被引用也有类似现象,期刊刊载论文的引文分布也有一个比较集中的核心区域,这就是著名的加菲尔德引文集中分散定律。"加菲尔德强调的引用关系本身带有天然的评价关系,是核心期刊具有评价功能的出发点。"①所以,无论是核心期刊,还是 CSSCI 来源期刊,都是以布氏和加氏理论作为其评价行为和评价产品理论基础的。由于布氏定律和加氏定律都是在对学术期刊进行分析和研究后得出的,故而在这样原理指导下的"三大核心"其实都是以学术期刊为对象的评价。

①姜晓辉:《核心期刊的评价功能与作用》,《澳门理工学报》(人文社会科学版)2012 年第 1 期。

必须指出的是,核心期刊与 CSSCI 来源期刊是有着本质区别的,这一点,稍后再分析。但至少在它们问世时,在世人眼里,并没有什么不同。不然,也不会有“三大核心”之称了。

其次,“三大核心”虽然都只是关于学术期刊的排名,但却为所有的评价定下了基调。学术评价种类繁多,诸如项目申报、绩效考核、头衔授予、职称评审等,都需要对单位、机构、团队或个人科研水平进行评价。其中的关键都是对评价对象所产出的学术论文价值的评价,因为学术成果最主要的表现形式就是公开发表的学术论文。评价论文不仅需要深入内容层面的专业分析和判断,还需要经过较长时段沉淀的历史检验,且见仁见智是常态,可见其难。而在同行评议失去公信力后,如何科学、公正且及时地评价论文,更加成为困扰人们的难题。核心期刊的问世为解决这一难题提供了替代方法,就是“以刊评文”。如同前些年采用的检测奶粉质量的方法一样,奶粉的质量标准应该是蛋白质的含量,但直接测定蛋白质含却是复杂和困难的,于是人们找到了间接办法,通过检测出奶粉中氮元素的含量换算出蛋白质含量,凭此判断奶粉的质量。所以,论文评价也可用间接的办法,即通过期刊过往的形式数据来“算出”其优劣,分出等级,再依据期刊的不同等级确定其后来发表的论文质量的高下。这样一来,有了期刊评价,更确切地说是有了期刊等级排行榜,就有了论文评价;有了论文评价,也就有了一切评价。故在所有评价中,期刊评价结果的基础地位谁也无法撼动。“三大核心”虽然只是期刊排行榜,却能左右绝大多数学术评价的奥秘就在于此。

再次,“三大核心”之所以成为所有评价的基础,还因为其制作者都声称他们所做的是客观的“定量评价”,故而具有天然的公正性和公平性。除了 CSSCI 制作者坚守了只依凭引文数据外,另两大“核心”制作者还进行了理论创新。“‘布拉德福文献离散定律’和‘加菲尔德引文集中定律’揭示了学科文献在期刊中的分布存在‘集中’和‘分散’的客观规律,为定量评价学术期刊奠定了理论基础。后人进一步研究证明,学科文献的被摘录、被转载、被阅读等多种特征在期刊中的分布都遵循集中和分散的规律。”①从而创造性地将对学科文献的摘录、转载、阅读等带有一定目的性行为的简单计量都视为与引文计量等量齐观的定量评价。如此一来,核心期刊评选指标的“丰富性”“全面性”就远远超过了加菲尔德只重引文而不及其他的单一性。他们还声称所有的数据不仅都是具有评价意义的,而且都是客观的,他们只是通过“科学”的算法,按其结果对所有的学术期刊进行排序,因而“公开”“公正”和“公平”是有保证的。在同行评议声誉每况愈下之时,这样量化的“客观评价”在 1990 年代的确曾经给学术界带来了科学合理评价的希望。

最后,“三大核心”都以一定的周期进行期刊排行榜的更新,每次更新其评价指标或方法都会有一定程度的调整,比如应行政权力的要求而增加某些指标,或为应对学术界的批评而调整某些指标的权重。这在“三大核心”说来是很正常的事,情况总是在不断变化的,学术研究和学术期刊也不会例外,故而排行榜也当常排常新。一般的排行榜对于被排行者来说,每一次更新,无非是前进了几名还是后退了几名的事,但“三大核心”的更新有点不一样,因为其制作的目的是为期刊划分等级,所以制作者公布的并不是对所有学术期刊的排名,而只是其中大约 20% 即被认定为“核心”或“来源”的期刊的排名,所以每次更新都是有进有出的,进出则意味着等级的变化。这样一来,对于个体的学术期刊来说,重要的问题首先是在不在这个榜单上,其次才是排在第几的问题。因此,尽管“三大核心”一次又一次地说明他们的评价是纯客观的,但在每次新版公布之前,都会有不少期刊主编前来公关或打探消息②;又因为刊物等级的变化直接影响到其所刊发论文的评价,故也会有已被某刊录用但尚未刊发的作者前来询问该刊是否还留在榜单上,以考虑是否需要撤稿;而在榜

①蔡蓉华、何峻:《论期刊评价之目的、方法和作用》,《澳门理工学报》(人文社会科学版),2012 年第 2 期。

②《中文核心期刊要目总览》主编之一蔡蓉华曾告诉《中国青年报》记者,每 4 年一次的核心期刊评审,都是一系列“公关”与“反公关”的过程。每到评审时节,就有不少学术期刊的主编们纷纷找到课题组,甚至动员同学、同乡、师友等各路人马说情(参见《“核心期刊”评选的背后:主编们频频公关》,《中国青年报》,2009 年 4 月 27 日)。

单公布之后,又都会掀起一场关于评价的热议,特别是那些被挤出榜单的期刊,有时还会表现得非常激烈,要过上一段时间后才会逐渐风平浪静。当下一个更新周期到来之时,这一切又会再演一遍,从而在客观上得以牢牢锁住期刊界和学术界的关注。

从以上分析不难看出,“三大核心”的评价结果既简单直观,又有广泛的适用性和关注度,而且还“客观”“公正”和“科学”,因此,它们被行政权力部门相中也就不奇怪了,因为具有这些特征的评价产品正是行政权力部门在学术资源分配和科研管理中最为需要和最好用的,故而“三大核心”才得以在如今学术研究的全领域、全过程中无处不在,由此也足见“三大核心”在我国的学术评价体系中的分量。其实,“三大核心”本身也常常会被冠以或自封为“评价体系”之名,意即它们是自成体系可以独立运作的。必须指出的是,行政权力之所以采信“三大核心”,最初其实与吃够了同行评议苦头的学术界对“三大核心”一定程度的认可是分不开的;而一旦为行政权力部门采信,在学术界当然更是通行无阻。只要看一下高校皆以期刊为本的相似的考核方式和各自制定的用于考核的期刊目录就可以清楚这一点,至今,这些目录无不以“三大核心”中的某一个为蓝本。

三

那么,以“三大核心”为基础的学术评价体制合理吗?时至今日,大概没人会给出肯定的答案,但到目前为止,所有对“三大核心”的批评和讨伐都集中在其评价方法或程序的不合理。其实,在回答“三大核心”的评价是否合理这一问题之前,有一个问题更值得追问,那就是,“三大核心”制作者们所做的是学术评价吗?要回答这个问题,我们不妨对“三大核心”的特点做进一步的分析。

首先,支撑“三大核心”的评价理论是否科学。

所谓“布氏定律”,说明的只是各学科论文在学术期刊中的分布规律。在专业分工和与此相应的专业期刊远不如今天这般发达,多数学术期刊学科边界尚不十分清晰的1930年代,布氏定律所揭示的论文分布规律的意义是不言而喻的,但在今天这样一个学术研究和学术期刊分工高度精细化的时代,布氏定律早已是明日黄花了。而不管布氏定律过时与否,其所揭示的充其量也只是不同学科的文献分布规律而不是不同质量的文献分布规律,与我们今天判断期刊学术质量高下的学术评价已没有直接关系了。再看“加氏定律”,它也不过说明得到较多引用的文献会集中在部分刊物上,当然,这与学术期刊的质量之间的确有一定的关系,但这个关系是复杂的和非线性的。因为引文虽然大多出自同行学者,似乎代表了同行对被引文献的某些意见,但引文的动机和目的却是多样的,比如有正面引用、负面引用,以及不具备实际意义的为了引用而引用等多种情况;而被引文献之于施引者研究的作用则更是复杂的,即使同样是正面引用,对施引者的作用也是不尽相同的。故而只有对施引者的动机、目的和被引文献之于施引者的贡献作出专业分析后才能对评价产生一定的意义。不经过专业分析的单纯数量统计是没有科学评价意义的,更何况不同学科甚或同一学科不同方向因引文习惯不同而导致的数据迥异,比如同样是引用同行学者的论文,经济学论文的引用数量可能是历史学论文的十倍以上,这岂是加权和算法所能对付的。即使强调引文评价作用的加菲尔德本人也是从不忘同时强调引文分析对引文评价的关键意义的。

至于“核心期刊”制作者对评价理论的“创新”,即引入复合指标,则至少存在着两方面的问题:其一,这些增加的指标是否具有评价意义?比如被视为重要指标的“文摘量”“文摘率”,已有实证研究证明,文摘的数据是不足以用来评价学术期刊质量的①。其二,即使这些新增的指标具有一定的评价意义,又如何判断其意义的大小?这是依据复合指标制作排行榜的评价所面临的普遍问题。

①参见王文军:《中国学术文摘:现状与展望——以“三大文摘”为中心的实证研究》,《清华大学学报》(哲学社会科学版),2013年第6期。该文论证了期刊被摘量与期刊学术质量并无正相关关系。

评价机构采取的应对方法无一例外,就是人工干预——加权。那么,如何加权?依据是什么?同样无一例外,评价机构的加权都是依据主观判断来进行,其实,加权原本就是一件极为主观的事。这样一来,所谓的客观评价就充满了主观色彩,对于排行榜的制作者来说,稍稍修改一下加权方案就能轻而易举地改变排序的结果。这也是每有新的期刊排行榜出炉之前,评价机构的门槛都要被前来公关的人踩平的原因所在。其实,这些登门的人与其说是希望得到某种“照顾”,不如说其中的大多数是担心评价机构数据采集和加权中的任意性而使自己的利益受损。如果说,量化评价的最大特点或优势是指标的“客观性”“公正性”,那么,随着在制作排行榜时引入复合指标和加权,“客观性”“公正性”也因失去了基础而大打折扣,甚至已不复存在。

可见,所谓评价理论特别是核心期刊制作者的诸多“创新”,并不那么靠谱。这也是“三大核心”中作为后来者的“C 刊”地位不断上升的原因之一,“C 刊”的制作者至少坚持了加氏定律的纯洁性,保持了数据的单一性,故其客观性和公开性远胜于另两个引入综合指标并加权运算的“核心”。如果被评价者必须受“核心”指挥的话,当然宁可选择公开性和公正性得到相对保障的“C 刊”。于是,在 2017 年初对 CSSCI 讨伐风潮中我们看到了一个以往从不曾见过的现象,就是与以往一面倒的讨伐不同,有著名期刊人站出来公开呼吁应公正地看待 CSSCI 之于学术研究、学术期刊和学术评价的作用①。

其次,导致“三大核心”作用不断外延的“以刊评文”是否科学。

今天,从学术界、学术期刊界到评价机构再到管理部门,对“以刊评文”均持否定态度,这应该说是近十多年来学术界和学术期刊界对以“三大核心”为代表的评价机构讨伐的一个可以看得见的成果,但这似乎并不妨碍“以刊评文”在现实中风行不变——几乎所有高校的评奖和考核仍然是在数“核心”论文的篇数。其实,“以刊评文”并非没有其自身的逻辑,这个逻辑是这样的:期刊之所以优秀是因为其所发表的论文优秀;所发论文之所以优秀,是因为期刊社或编辑部执行了严格的审稿标准,将质量低劣的来稿拒之于门外。当然,也就可以反过来推论:发表在优秀期刊上的论文都是经过了严格审稿的,故而都是优秀的,而“三大核心”榜单上的期刊都是优秀的,能在这些期刊上发表的论文一定都是经过了严格审稿的,当然也就是优秀的。如果“三大核心”对期刊的遴选能基于被评选期刊的所有论文,证明入选期刊发表的每篇论文学术质量都是有保障的,那么,这样的逻辑演绎亦即“以刊评文”就不会有太大的问题。

但是,现实中“以刊评文”的结果却十分荒唐,连“三大核心”的制作者们也不得不承认这一点。那么,问题只可能出在评刊,即“三大核心”的遴选标准上。秘密在于:其一,“三大核心”各自评选指标的数值很容易被少数乃至个别论文的极端值所左右。在“三大核心”榜单遴选时,并不针对期刊发表的所有论文,即不要求入选期刊所刊发的每一篇论文都具有指标所规定的足够数据值。事实上,“三大核心”也都没有针对期刊所有论文的指标存在,故而只要其中部分论文具备一定数据值,其余文章,则可完全无视。其二,“三大核心”各自评选指标只涉及形式数据而不涉及内容。这就如同通过检测氮元素的数量来评价奶粉质量一样,为数据造假留下了足够的空间,不法厂商在奶粉中掺入氮元素丰富却非蛋白质的三聚氰胺即可造成高蛋白质的假象,学术界的“三聚氰胺”——人为造假的数据同样可混迹其间,造假的方式则是五花八门,层出不穷,令人目不暇接。

正是评价机构极力将刊物的部分论文数据等同于全刊数据,又难以准确地鉴别数据的真伪,更不分析数据本身的意义,才造成了“以刊评文”结果的荒唐和恶果的蔓延。不管是逻辑推演还是事实判断,“以刊评文”都首先错在评刊,是评刊出了问题才导致“以刊评文”跟着出问题。而当错误的评刊成了“指挥棒”,相当多的期刊社或编辑部就没有必要坚持统一的用稿质量标准。至于评价机构也在高调反对“以刊评文”,这是因为他们明白自己根本没有能力基于论文的质量来评刊,按

①参见仲伟民、桑海:《如何客观评价 CSSCI》,《澳门理工学报》(人文社会科学版),2017 年第 3 期。

照他们的方法所评出的“核心”中并非所有论文的质量都有保证,所以不能“以刊评文”。当然,他们也不会不明白论文评价之难,“以刊评文”根本就没有替代品,所以高喊不能“以刊评文”,既可为他们开脱错误评刊所应负的责任,又不会危及自身的利益。

再次,“三大核心”的所谓“客观”的定量评价是否科学。

所谓定量评价,一般指的是通过对评价对象相关数据的采集、分析、运算,得出相关结果的方法。作为方法,定量评价或量化分析具有广泛的适用性,早已成为包括人文学科在内各学科普遍采用的分析方法。因此,问题不在于可不可以用,而在于谁来用和怎样用。

定量评价的基础是数据,但是,并不是所有的可以用于评价的信息特别是涉及评价对象内容层面的信息都可以转化为被纳入“算法”用作运算比较的数据的,即使可转换为评价数据的信息也无法避免数据转换过程中出现的某些重要信息的丢失、扭曲和偏差。因此,数据虽然客观却难免片面和失真,对于以学术研究成果或学术期刊为对象的复杂评价,单纯的定量评价是无法胜任的,而只能是工具性和辅助性的。这就决定了仅依靠对期刊某些形式数据的运算即得出结果的评价机构所做的必定是变形的和不完整的评价,谈不上客观,更谈不上公正和公平。同样重要的是,即使是作为工具性和辅助性的定量评价,也必有一个绝对不可缺损的程序,即对数据的取舍、分析和价值判断,这是决定其“算法”是否科学和专业的前提。对于成果充满个性的人文社会科学研究和期刊而言,数据所代表的意义需要非常专业的分析和判断,故定量评价也只有同行专家才有可能胜任。评价机构所缺的正是专业水准,他们总希望用“算法”来替代专业分析,但至少到目前为止,再复杂的算法也不可能取代专业分析,何况他们所谓的“算法”只是极为简单而随意的加权而已。换言之,没有专业水准,就不可能有科学的“算法”,这就决定了评价机构不可能作出科学和专业的评价,即使单纯的定量评价,也不是其所能胜任的。

最后,评价机构是否为合适的评价主体。

尽管“三大核心”中两家只生产排行榜而不开放其数据的制作者始终声称自己是合适的甚至权威的评价主体,但这样的主体身份是经不起推敲的。因为囿于专业所限,“三大核心”制作者都不可能从事对于其挖掘和整理的数据的专业分析,而缺失了这一学术评价必不可少的关键环节,即使仅做定量评价,他们也不可能是合适的主体。“三大核心”制作者们所做的只是某些期刊数据的定向挖掘和整理工作,而且在“三大核心”中,除了 CSSCI 开放其数据库以外,另两大“核心”那多达十几个乃至数十个指标的原始数据的挖掘过程和结果从未开放过,而《中文核心期刊要目总览》的制作者居然已放弃了原始数据的挖掘和采集。即使“三大核心”都能开放其数据,充其量也只是为专业的评价做了某种准备而已,而决不是真正的评价。

近年来,CSSCI 的制作者已公开放弃了作为评价主体的身份,也不再将其来源期刊目录视为评价成果,而其主打产品则是 CSSCI(人文社会科学引文数据库),这个数据库的开放存在才是其能够公开放弃评价主体身份的底气之所在。这使得 CSSCI 制作者彰显出了与另两大“核心”制作者的根本不同:后者不仅始终坚持他们是合适的评价主体,而且他们的评价产品只有期刊排行榜而没有开放的数据库,故其排行榜即使真的按某种“算法”而得出的,学术界和学术期刊界也无法对其进行必要的监督,其同行也无法重复其运算过程,所谓的科学性也就可想而知了。其实,“三大核心”的三个期刊榜单之于学术研究和学术评价都没什么意义,如果有,也只能是负面的。但需要指出的是,他们所挖掘和整理的数据是有意义的,特别是 CSSCI,它并不是什么“核心期刊”,而是引文数据库,将其作为数据源的“来源期刊目录”等同于“核心期刊目录”,是对 CSSCI 的最大误读。关于这一点,稍后我们将作进一步的分析。

当我们弄清楚了“三大核心”制作者并不具备评价主体的身份,其所做的也不可能是真正的学术评价,“三大核心”(即期刊排行榜)也算不上是所谓的评价结果之后,以“三大核心”为基础的学术评价体制是否合理的问题也就不言自明了。

四

以上我们分析了“三大核心”为代表的量化评价的依据和缺陷,这样的量化评价的根本缺陷就在于外行评价内行。但是,这个外行之所以能够获得对内行的评价权,不仅因为他们获得了行政权力的采信,还因为“内行”在评价方面的无能和堕落而失去了学术公信力。诚如张耀铭教授所言:“公信力是权威性资源之一,一旦受到损伤,便会造成较大面积的信任危机。”①正是学术界中的许多人对同行评议的失望,才有了1990年代对量化评价的呼唤,从而使得评价机构乘虚而入。所谓请神容易送神难,当学术界发现这个他们寄予很大期望的量化评价远没有设想中的那么美好,甚至进一步毒化了学术生态时,“三大核心”早已成尾大不掉之势。对于迅速崛起的评价机构,行政权力则经历了从对“三大核心”运用的得心应手到发现其问题却投鼠忌器的过程。导致学术界和行政权力对“三大核心”态度变化的直接原因是频发的学术不端现象。

在学术评价中引入“三大核心”二十多年后的今天,学术不端已是司空见惯的现象,以至于一般的“炒冷饭”和重复发表,甚至少量隐晦的抄袭剽窃已引不起人们的兴趣,而每隔三两年总会有一两件匪夷所思的学术不端事件曝光,诸如院士造假、校长剽窃之类,才得以引起公众的广泛关注。关注的结果之一就是将这样的事件与学术评价联系起来,视不合理的评价机制为其根源,当然也就少不了对评价机构的一次新的讨伐。

最近一起这样的事件当属2018年10月24日《中国青年报》一篇题为《青年长江学者与她“404”的论文》的报道所曝光的“青年长江学者”梁教授从各家期刊数据库一口气撤下了自己百余篇论文事件。在这之前,梁教授无疑“是同龄人中的佼佼者”,她几乎拿下了这个年龄的大学教师能获得的所有荣誉。撤稿的原因据梁教授自己说是因为早期论文学术质量不高和不符合学术规范,但却不能不让人联想到这些被撤下的论文更可能与某些学术不端行为相关,故而不出意外地激起了人们的愤慨,而梁教授面对记者说出的一句“我这条路有多难你知道吗”②,再一次地使人联想到学术评价机制之“逼良为娼”,引发了许多人的感慨。梁教授的成功之路显然与她学术成果数量的丰硕是分不开的,我们暂且将她是否涉嫌学术不端而撤稿这一疑问置于一边,即使如她所说撤稿是因为论文的质量不高和不符合学术规范是真的,那么,她无疑证实了自己就是以量制胜的典范。“量”何以制胜?当然靠的是以“量”为标准来评判学术优劣的评价。梁教授的成功似乎再次向世人昭示了迎合排行榜的“意义”——学术利益的获取,从而折射出了学术评价机制的不合理。然而,梁教授并非一个孤例,她只是近十多年来迎合学术评价的受益者之一,她的成功和她的悲剧都只不过是比别人多走了一步而已。她当然不是第一个也决不会是最后一个因迎合评价而获益和同样因迎合评价而乐极生悲的青年教师。近年来,梁教授式的成功和悲剧已成为讨伐学术评价机构的有力武器,使得评价机构平添了许多压力,不得不努力改变和重塑自己的形象。

更能感受到学术不端行为频发而带来压力的是行政权力部门,毕竟评价机构的所作所为在很大程度上是出于现行科研体制顺畅运行的需要而由行政权力默许甚至明许的,因为决定评价机制的正是行政权力部门。对于“三大核心”,行政权力的心态是颇为复杂的。排行榜将复杂的学术评价简单化,使行政权力运用起来得心应手固然是他们不愿放弃这样的评价的原因,比如原新闻出版总署2010年下发的《全国报纸期刊出版质量综合评估指标体系(试行)》就是典型的“三大核心”的翻版,而找不到如此好用的替代品则是另一个原因。但是,来自学术界和学术期刊界对以“三大核心”为代表的学术评价机构及其产品的讨伐,以及将学术不端的根源指向现实中的学术评价机制

①张耀铭:《重建学术期刊的公信力和权威性》,《澳门理工学报》(人文社会科学版),2015年第2期。

②《青年长江学者与她“404”的论文》,《中国青年报》,2018年10月24日,第10版。

甚至科研体制,使得行政权力部门对“三大核心”的采信不仅留有很大的余地,而且对“三大核心”在学术评价中的运用可能造成的混乱保持了一定的警惕,也意识到改革不合理的学术评价机制之必要,所以也一直倡导着要建立科学合理的学术评价机制。时至2010年代中期,对评价问题的重视已由具体的科研管理部门如教育部、科技部等上升到了最高行政权力部门。

2016年1月,国务院办公厅发布了《关于优化学术环境的指导意见》,该文件开宗明义即指出:“目前我国支持创新的学术氛围还不够浓厚,仍然存在科学研究自律规范不足、学术不端行为时有发生、学术活动受外部干预过多、学术评价体系和导向机制不完善等问题。”针对学术评价领域各类排行榜泛滥的情况,该文件在“任务要求”部分特别指出:“优化学术民主环境,营造浓厚学术氛围……鼓励开展健康的学术批评,发挥小同行评议和第三方评价的作用。科学合理使用评价结果,不能以各类学术排名代替学术评价,避免学术评价结果与利益分配过度关联。”①这是笔者所见第一个明确提出“不能以学术排名代替学术评价、避免学术评价结果与利益分配过度关联”的国务院文件。该文件的发布,说明在关于学术评价的顶层设计中,已意识到必须警惕排行榜的负面作用。然而,该文件发布后,学术界却波澜不兴,在高校和科研院所的各类评价中,以排名代替学术评价、学术评价与利益分配紧密关联的现象仍然普遍存在。正因为如此,2017年初对“CSSCI”的讨伐仍然如期而至。与学术界不同,对排名就是评价的否定,评价机构显然要敏感得多,因为他们手中所握有的评价权力是由行政权力部门所赋予的,来自行政权力部门特别是最高权力部门对学术评价态度哪怕任何细微的改变,都足以对评价机构的前途造成决定性的影响。这一点,下文再予分析。

就在国务院办公厅《关于优化学术环境的指导意见》发布两年多后的2018年7月,中共中央办公厅、国务院办公厅印发了《关于深化项目评审、人才评价、机构评估改革的意见》,这是近年来“中央两办”首次共同就学术评价问题发布专门的指导意见,再次昭示了顶层设计对学术评价的重视和改革评价机制的决心。该文件从“优化科研项目评审管理”“改进科技人才评价方式”“完善科研机构评估制度”“加强监督评估和科研诚信体系建设”等方面对学术评价提出了系统性的改进意见。其中,关于“改进科技人才评价方式”部分特别提出:“科学设立人才评价指标。突出品德、能力、业绩导向,克服唯论文、唯职称、唯学历、唯奖项倾向,推行代表作评价制度,注重标志性成果的质量、贡献、影响。把学科领域活跃度和影响力、重要学术组织或期刊任职、研发成果原创性、成果转化效益、科技服务满意度等作为重要评价指标。在对社会公益性研究、应用技术开发等类型科研人才的评价中,SCI(科学引文索引)和核心期刊论文发表数量、论文引用榜单和影响因子排名等仅作为评价参考。……注重发挥同行评议机制在人才评价过程中的作用。”②“克服唯论文、唯职称、唯学历、唯奖项倾向”以及“核心期刊论文发表数量、论文引用榜单和影响因子排名等仅作为评价参考”的提出,矛头所向,直指单纯的量化评价,而“三大核心”排行榜正是这样的产品。

反“四唯”的口号一经提出即得到了广泛响应,而且与以往不同的是,此次的反“四唯”并没有仅仅停留在口头上,而是有实际行动的。国务院办公厅专门发布了《关于优化科研管理提升科研绩效若干措施的通知》,该文件特别指示:“开展‘唯论文、唯职称、唯学历’问题集中清理……对项目、人才、学科、基地等科技评价活动中涉及简单量化的做法进行清理,建立以创新质量和贡献为导向的绩效评价体系,准确评价科研成果的科学价值、技术价值、经济价值、社会价值、文化价值。”③

---

①国务院办公厅:《关于优化学术环境的指导意见》(2015年12月29日),国办发〔2015〕94号,中华人民共和国中央人民政府官网,http://www.gov.cn/zhengce/content/2016-01/13/content_10591.htm。

②中共中央办公厅、国务院办公厅:《关于深化项目评审、人才评价、机构评估改革的意见》,中华人民共和国中央人民政府官网,http://www.gov.cn/zhengce/2018-07/03/content_5303251.htm。

③国务院办公厅:《关于优化科研管理提升科研绩效若干措施的通知》(2018年7月18日)国发〔2018〕25号,中华人民共和国中央人民政府官网,http://www.gov.cn/zhengce/content/2018-07/24/content_5308787.htm

接着,科技部、教育部、人力资源社会保障部、中科院和中国工程院联合发布了《关于开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动的通知》,从“任务依据”“清理范围”“清理方向”“组织实施”等几个方面对各部委清理工作作出了部署。一场颇具声势的“四唯 ”专项清理工作就此展开。上述各部院及自然科学基金委、中国科协以及行业主管部门和地方相关部门都划出了重点清理的范围,涵盖了各种计划项目、人才项目、基地建设、机构评估、成果奖励、院士选拔、职称评审、绩效考核等几乎所有需要或涉及评价的方面,要将这方方面面涉及“四唯”的做法进行清理。具体清理方向是:“1. 对部门和单位政策文件中涉及‘四唯’的规定进行修改;对本部门和单位牵头执行的法律和行政法规中涉及‘四唯’的规定,提出修改建议。2. 对各类考核评价条件和指标中涉及‘四唯’的内容进行调整,具体表现形式包括但不限于评价指标体系、评价手册、评审细则等。3. 对有关管理信息系统和工作表格中涉及‘四唯’的内容进行修改。”①如此明确而具体的部署可谓史无前例,对学术界、学术期刊界特别是评价机构的震动可想而知。

这样的专项行动将会带来怎样的后果?《人民日报》发表时评作了乐观估计:“专项行动的开展,有望形成合力,转变科研中存在着的不当作风。”②笔者当然也希望能够早日呈现这样的好结果。然而,冰冻三尺非一日之寒,对于“四唯”的批判并非始于今日,近十多年来,学术界的有识之士们一再呼吁应该改革以“四唯”或其中的某“唯”为标志的不合理的学术评价机制,也得到了行政管理部门的回应,可结果又如何呢? 当然,这一次的反“四唯”专项行动的确与以往不同,最大的不同在于,一是由最高层发动;二是有具体的清理方向、范围和措施,但我们必须看到,清理“四唯”主要还属于“破”的范畴,而“破”字当头,“立”是否就在其中了呢? 这就要回答“立”什么,如何“立”的问题,这才是问题的关键。

五

所谓“立”什么、如何“立”的问题,实际上就是如何为以“三大核心”为代表排行榜寻找到合适替代品的问题。近年来,从管理者到学术界以及学术期刊界对学术评价问题多有思考,现行评价之不合理是共同的判断,甚至连评价机构都不否认这一点。但不合理在何处、造成不合理的原因是什么,各方的看法却不尽相同,所以如何改进也是众说纷纭。总的说来,改进的路径设定大多是以下两个之一,即或者主张改进评价方法,以使其更加科学,比如提倡“定量评价与定性评价相结合”;或者主张改变评价主体,使评价向同行评议回归,比如启动“代表作评价制度”。对这两者,很多人都抱有很大的希望。在笔者看来,设想都很美好,但可行性如何,恐怕还得打个问号。

先看“定量评价与定性评价相结合”。

关于“两结合”的好处,一般的解释是让这两种各有所长、也各有缺陷的评价互补,以使评价变得相对全面和公正。在笔者看来,这个已被普遍接受的替代方案貌似公允,其实很容易滑向似是而非。这是因为,如果仅仅将定性和定量视为两种评价方法,同一评价主体可以根据评价目的和内容的需要而灵活地加以运用时,“两结合”与所谓方法的多元是一个意思,当然是可以的;但当定性和定量成为两类不同主体独立的评价,而“两结合”只是将它们各自的结果简单地叠加时,评价结果的科学性和公正性就很值得怀疑了。而今天各界所主张的“两结合”,恰恰多是后者而鲜见前者。

---

①科技部、教育部、人力资源社会保障部、中科院和中国工程院联合发布《关于开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动的通知》(2018 年 10 月 15 日),国科发政〔2018〕210 号,科技部官网,http://www.most.gov.cn/tztg/201810/t20181023_142389.htm。

②沈彬:《人民时评:校正科研激励的指挥棒》,2018 年 10 月 31 日,人民网,http://theory.people.com.cn/n1/2018/1031/c40531-30372591.html。

之所以说当“两结合”成为定性和定量两种评价结果的简单叠加无法保证评价结果的科学性和公正性,是因为当我们判断一个评价是否合理时,需要从以下几个方面去分析。

首先,我们要看谁是评价主体。“两结合”恰恰模糊了谁是评价主体这一关键问题,似乎学术共同体和评价机构是对等的评价主体。自从“三大核心”制作者推出其“定量评价”的排行榜之后,经过刻意渲染,定量评价成了评价机构的专利,而各学科专家只会做定性评价。他们全然不顾早在评价机构产生之前,定量评价作为一种基本的研究方法就已被包括人文学科在内的各学科专家所普遍使用这一事实(比如在史学术界“定量史学”就成功地使用了定量分析和评价的方法,更不用说在经济管理等社会科学各学科了)。于是乎,在学术评价领域就出现了一个奇特的现象,不同的方法只能由不同的人来使用,而不能由同一人或同一类人来使用。评价机构借此以“定量评价”的唯一适格使用者而跻身评价主体的行列,从而得以与被描述为只会做定性不会做定量评价的各学科专家(学术共同体)分庭抗礼。这样一来,原本归属于学术共同体的评价就分裂为评价机构的定量评价和学术共同体的定性评价这两类互不统属、完全独立的评价,学术评价就此分裂。评价的分裂如今已成不争的事实。评价的分裂之于学术研究和学术期刊的后果是相当严重的,因为分裂后的“定性”和“定量”都成了残缺不全的评价。在当前令人忧虑的学术环境下,缺乏自律的定性评价(同行评议)往往又呈现出不客观、不公正、不科学,其对学者个人利益的伤害有时甚至超过了评价机构的排行榜,两相比较,人们宁可相信排行榜。这就使得学术评价几乎全由评价机构说了算,成了他们的一统天下。“两结合”正是针对这种情况而提出的应对方案。我们知道,在评价问题上,行政权力部门、评价机构和学术界是有着不同的利益诉求的,有时甚至是相互对立的,但在“两结合”提出后,三方却出奇地一致拥护。个中原因是什么?正因为这个方案并没有明确谁是评价主体这一关键问题,只要打出“两结合”的幌子,谁都能当评价主体。据此,行政权力部门评价组织者的身份不会改变;评价机构虽然不可避免地会出让某些已到手的权力,但仍可通过具体操作予以弥补,评价主体身份这一要害不会改变;而学术界似乎也找到了重返评价领域的路径。于是,三方的利益和诉求都在一定程度上得到了满足,但是,谁是评价主体的根本问题并没有得到解决,评价的分裂不仅依然如故,而且还因为“两结合”的提出而合法化了。这样的“两结合”又怎能走向科学?

其次,我们要看运用的是怎样的评价方法。评价方法总是与评价目的和内容相关的,不同的评价目的决定了不同的评价内容,不同的评价内容决定了必须选择不同的评价方法,只要评价目的和内容不同,所采取的方法也就各异,不涉及目的和内容谈方法都是没有意义的。这应该是常识。至于应该使用什么方法,完全取决于评价主体根据目的和内容的需要进行的选取。仅就方法层面来说,不管选择定量还是定性,有一点是肯定的,就是评价者必须具备专业分析能力,否则,定性评价就难免望文生义,而定量评价则会沦为数字游戏。可见,除了评价目的和内容外,评价方法还是离不开评价主体,在不明确评价主体的情况下,“两结合”连评价方法的改进都难做到,只能退而求其次,寄希望于以两类评价结果的糅合来替代主体的统一和方法的改进,但因分裂而残缺的评价并不会因此而真正统一起来,残缺依旧。这样的残缺,并不是缺损了部分正确的结论,余下的还是正确的,而是因主体的能力或方法的残缺导致了所有结论都可能是变形的或错误的,我们怎能指望将这样的两种结果叠加就可以得出科学的结果来呢?

再次,我们要看评价的程序是如何设计的。评价的程序如何设计直接关系到评价的公正与否,程序公正是评价公正的前提,这是不言而喻的。通过评价程序,我们可以看到谁是评价的组织者,谁是评价主体,谁是评价的参与者,主体和参与者是如何分工的,以及经过了哪些步骤而得出了评价结论,可见评价程序仍然与评价主体相关。迄今为止,我们还没有看到“两结合”提倡者如何设计评价者身份和如何设计“两结合”程序的理论探讨,但却不乏一些可供我们分析的实例。已有的“两结合”实例,评价的组织者大多不外乎两种,一是行政部门,二是评价机构。即使退一步说,“两结合”就是两类评价的糅合,那么在程序设计上,至少也应该让两类主体具有对等的地位。实际情

况又是如何呢?先看评价机构组织的“两结合”评价的程序设计。目前评价机构大多号称其排行榜是“定量评价与定性评价相结合”的产物,因为他们引入了学科或期刊专家参与排行榜的制作。评价程序是如何安排专家评刊的呢?无非是让成百上千的专家在一张表格上凭印象对成百上千的期刊进行打分。事实上,对于如此数量级的期刊,专家熟悉的只是其中的少数乃至极少数,这样的打分又有多少参考价值?而这些专家所打的分在排行榜制作中到底起了什么作用,恐怕连这些专家也不知晓,最终的算法和评价结果,这些专家更是没有决定权的。在这样的评价程序中,专家最多也只参与了一个环节而不是全过程,哪有什么主体地位?起决定作用的仍是评价机构。再看一些行政主管部门组织的评价,类似的所谓“两结合”的程序设计也比比皆是。所谓专家参与,不过是按定量评价的“指标体系”逐项打分而已,而对这些指标体系是否合理,最终结果如何,专家一般无权置喙。在这样的“两结合”程序设计中,专家只是起到了“打分机器”的作用,谈何主体地位?如此的“两结合”程序设计,又有多少科学性可言?

最后,我们还要看数据在评价中是如何发挥作用的。作为分析工具的数据处理在评价中的作用是十分重要的。如前所述,数据如何发挥其应有的作用,取决于采集、整理、分析、运算、解释的过程是否科学,关键还在于进行数据处理的人是否专业,故而只有同行专家才是最合适的数据处理人选。因此,在“两结合”评价中,数据能否发挥应有的作用,还是离不开主体的设定。但在评价分裂情况下的所谓“两结合”评价中,由于评价机构已获得了定量评价的专利,也就意味着不仅所有的评价数据都由其提供,而且还意味着提供什么样的数据、对这些数据进行怎样的整理、分析、运算和解释,其权利都操于评价机构之手,被“结合”的专家们既无法审核其原始数据,也无法对数据作出分析,更无法干预其算法,除了为评价机构的“两结合”背书以外,还能做什么?比如“三大核心”中两个核心期刊的评选,都号称吸纳了成百上千专家的参与,但专家们是审核了数据,还是提供了对数据的专业分析抑或是干预了算法?当专家们参与进去的时候,评价机构早已“算”出了排行榜,专家们充其量也就是凭主观印象对结果做些微调而已。我们又如何指望这样的“两结合”能真正发挥数据以及专家在数据分析方面的作用?

从以上分析不难看出,导致“两结合”滑向似是而非的是评价的分裂,不解决评价主体是谁的问题而只是依靠糅合两类评价的办法是不可能走出评价困境的。在我们明确了学术评价的主体只能是学术共同体后,就可以发现科学的评价并不在于定性评价与定量评价这两类评价的结合,而在于让定性评价和定量评价都回归到一般的评价方法,由评价主体根据不同的评价目的来选择和确定最合适的方法。这个方法,可以是定性的,也可以是定量的,更多的时候则是多元的。如同学术研究一样,研究者可以根据不同的研究目的和所掌握的不同资料,采用不同的方法。方法可以多种多样,但不同的方法只能由同一个主体来统摄,不应分别成为不同主体的专利,而有能力统摄评价方法的主体只能是学术共同体。但是,让学术共同体回归评价主体的地位却有着种种困难,而“代表作评价制度”的提出,就是要为同行评议的回归架设一道桥梁。这条路能走得通吗?

六

再看“代表作评价制度”。

所谓“代表作评价制度”大致是指通过对被评价者最具代表性的科研成果按一定的程序、方法进行评审,得出关于被评价者学术水平的结论的学术评价制度。近年来部分重点大学在职称评审等涉及评价的事务中高调推行这一制度,令人有耳目一新之感,赢得了很多赞扬,也得到了教育部的首肯,教育部近年正在力推这一制度,而在“中央两办”《关于深化项目评审、人才评价、机构评估改革的意见》中,也明确提出了“推行代表作评价制度”。

之所以如此,是因为在学术界已风行多年的所谓定量评价——“以刊评文”导致的“以量取胜”

实在走得太远了。学术界普遍将时下种种学术乱象归因于定量评价的滥用,而推行“代表作评价制度”的目的就是要改变定量评价重“量”轻“质”的偏颇,重新确立以“质”为主的评价标准。既然是新推出的一项科研管理的重要制度,当然就需要有一系列相应的标准、程序等设计,以使该制度得以落实。部分学者也撰文参与了相关规则与程序设计的讨论①。但是,伴随着“代表作评价制度”的推行,在多方赞扬声中,质疑的声音也几乎同时出现了。质疑的声音虽然零星而不成气候,但仍然不容忽视,因为其所质疑的并不是以“质”为主的评判标准,而是如何具备可操作性,亦即如何保证“代表作评价制度”下的评价能够公正和公平②。

应该如何看待“代表作评价制度”?它能否取代运用排行榜的“以刊评文”?在笔者看来,“代表作评价制度”的实质是向以学术共同体为评价主体的回归,用同行评议代替排行榜。这无疑是学术评价走向科学和公正的必由之路。“代表作评价制度”,应该说是在领教了扭曲的量化评价的种种弊端和伤害后,学术界近年来寄予较多希望的一种评价制度,也可视为行政权力部门回应学术界的呼声而力推的一种取代“以刊评文”的重要制度。但是,提出这一制度设计与将这一设计落到实处之间是有一条鸿沟需要跨越的。如果我们不太健忘的话,当会记得在评价机构出现之前,同行评议一直是学术界通行的评价制度,这一制度之所以在 1990 年代以来被逐渐颠覆,学术共同体之所以几乎被驱逐出了评价领域,盖因这一制度得以公平、公正实施的条件已渐渐地丧失了。

要确保以学术共同体为评价主体的同行评议的科学、公正是需要具备一定的条件的,至少评价主体必须满足以下两个条件:其一是专业性;其二是自律性。只有同时具备了这两个条件,同行评议才可望得出令人信服的结论,否则也许连评价机构的排行榜都不如。

专业性对于学术共同体似乎不该是一个问题,要保证专业性,无非是在选择评价专家时注重其专业背景,一般说来就是“小同行”更有发言权。但是,专业性永远是有限度的。

20 多年前的“索卡尔事件”人们一定不会忘记。1996 年,美国纽约大学物理系教授索卡尔(A. Sokal)将其一篇题为《超越界线:走向量子引力的超形式的解释学》的“诈文”投给了著名的《社会文本》(*Social Text*),在这篇文章中,“索卡尔刻意模仿那些科学文人的语气,煞有介事地使用后现代主义的种种词汇,令人眼花缭乱地从量子力学,广义相对论,谈到微分拓扑,非线性,超弦和混沌,举出了一些似是而非的‘论据’,还杜撰了所谓‘形态发生场’的‘最新’物理学理论,通过这样的生拼硬凑胡说八道来论证‘量子引力是一种典型的后现代科学’,是‘超越界限’地发展出来的一种‘具有解放意义的科学’。从而得出结论说,‘后现代科学的内容和方法论,为进步的政治纲领提供了强有力的思想上的支持。’索卡尔在这篇文章里还引用了前辈物理学家玻尔(N. Bohr)和海森伯(W. Heisenberg)的言论作为依据,更加迷惑了刊物的几位总编辑,使他们看不出来这竟会是一场恶作剧。”该文发表后,“索卡尔立刻又写了一篇《曝光——一名物理学家的文化研究》,发表在同年晚些时的法语刊物《大众语言》(*Lingua Franca*)上。他说自己写了那篇文章来嘲笑科学文化研究者们,是因为:‘最好的武器就是讽刺,对那些最顽固对象的批评就是让他们自己打自己的耳光。’”③这就是轰动一时的“索卡尔事件”,由此引发了一场被称为“科学大战”(science wars)的激烈辩论。对于这一事件,人们可以从不同的角度进行解释,在此,仅这一事件中“诈文”居然得以骗

①此类文章的共同特点是,在论述代表作评价的意义与必要的同时,分析实行代表作评价所面临的现实困难,开列必须具备的条件,并设计代表作评价的程序。参见陈云良、罗蓉蓉:《“学术代表作”制度的实施条件和程序安排》,《现代大学教育》,2014 年第 1 期;姜春林、赵宇航:《代表作评价:探索之路与完善之策》,《甘肃社会科学》,2016 年第 3 期;王兵:《科研代表作评价工作平台的功能分析》,《中国高校科技》2017 年增刊;等。

②参见李涛:《高校学术评价不宜简单采用“代表作”制》,《中国社会科学报》,2012 年 10 月 26 日;石晶、李晓彤:《热与冷:高校学术代表作评价制度的思考》,《甘肃社会科学》,2014 年第 6 期;等。

③关洪:《“索卡尔事件”和科学家的介入——科学家与反科学思潮的一场较量》,《科学技术与辩证法》,2003 年第 5 期。

过著名期刊的专业编辑一端,就让我们不得不对同行评议中的专业判断打一个问号。

如果说这只是一个孤立事件,且《社会文本》当时尚未实行严格的同行匿名评审,还不足以说明同行评议专业性方面的问题,那么,在20多年后的2018年,被称为“索卡尔事件2.0版”的学术恶作剧的发生就再次提醒我们应看到同行评议的局限性。“波特兰州立大学的助理教授Peter Boghossian和他的几位同伴公开揭示了他们精心筹备一年之久的学术恶作剧——他们把伪造的二十篇论文寄给人文社科领域的知名期刊,结果竟然有七篇通过重重审议得以发表……这些数据可疑、结论荒谬的论文能够轻易蒙混过关的事实再一次激发了大众对于人文社科部分领域严肃性与学术价值的嘲讽和质疑。”①尽管索卡尔事件2.0版针对的主要目标不在同行评议,但同样可以用来说明同行评议的专业性并非无可挑剔。事件所涉及的期刊,无一不是实行严格的同行专家匿名审稿制的。索卡尔事件2.0版虽然仍然是一个恶作剧事件,但在此事件中竟然有三分之一多的著名期刊中枪,怎不令人震惊!匿名审稿制无疑是同行评议之一种,而在所有种类的同行评议中,匿名审稿是受到其他因素干扰最少的。不能不说在学科分化和专业分工日益精细的当下,同行评议的专业性正在受到前所未有的挑战。

与专业性相比,同行评议所要求的自律性所遭遇的问题更为严峻。如前所述,学术共同体被“三大核心”取代而逐步退出学术评价是从1990年代开始,造成这一局面的主要原因是同行评议的自律性遭到了大范围的质疑而不在其专业性的缺陷,自律性的失去与学术生态的恶化是分不开的,如李剑鸣教授所言:在“权力支配,人情主导,标准缺失”的大环境下,“三者只要居其一,都会使学术评价的意义受到严重损害”②。那么,今天的学术生态有根本性改变了吗?学界中人大概都会给出否定性的回答。因为学术生态恶化所导致的学术不端现象不是减少而是增多了,其程度不是减轻了而是更严重了,甚至到了公众见怪不怪,若没有挑战底线的事情发生都不算个事的地步。

最近的一起轰动全国的挑战底线的事当数北京电影学院翟博士学术不端事件,该事件起因于这位博士在网上晒出北京大学光华管理学院博士后录取通知书,很快有网友扒出了该博士竟然不晓知网为何物的视频,一位博士怎么可能不识知网?这实在是挑战了所有皓首穷经苦读学位的学子们的底线,于是,围绕该博士的学位何以获得的一系列猫腻都相继被扒了出来。对于这个引起公愤的事件,涉事的北京电影学院和北京大学及时作出了回应和处理,不出所料地取消了涉事者的博士头衔和博士后入站资格。舆论界对翟博士的学术不端行为有大量的报道和声讨文字③,这里只想指出,翟博士的学术不端行为固然值得声讨,但同行评议的声誉也因此降到了冰点更值得深思和警惕。如果说,前述梁教授因钻了量化评价的空子而获益,那么,翟博士的“成功”与同行评议的堕落则是分不开的。无论是博士学位的授予还是博士后录取资格的授予,都绕不过同行评议,前者需要通过严格的答辩,后者需要通过资格审查和面试,这些都可归为同行评议,而且是小同行评议。有关授予单位既然用了短短几天或十几天的时间就可以调查清楚并明确作出结论,显然弄清该博士的学术水平和学术研究的价值是否与其获得的学位和资格相称、是否存在学术不端行为并不困难,为何此前由同行专家组成的答辩委员会和录取机构的考核把关形同虚设?如果说该事件只是个案,并不能说明问题的普遍性,那么,近年来那么多日理万机的官员纷纷在职通过学术论文答辩而获得博士学位,难道他们都是天赋过人的学术奇才?

如果说匿名审稿制是对同行评议专业性的考验,那么,学位论文答辩则是对同行评议自律性的考验。学位论文答辩是典型的“代表作评价制度”下的评价,学位论文应该说是学位申请者至少经三年苦读而拿出的最具代表性的成果,答辩委员的选择也有严格的专业资格和标准设定,但这样标

①王晨阳:《学术恶作剧成功发表:人文社科领域被后现代话语败坏了吗?》,澎湃新闻,2018年10月15日。

②李剑鸣:《自律的学术共同体与合理的学术评价》,《清华大学学报》(哲学社会科学版),2014年第4期。

③关于这一事件的来龙去脉网上有大量报道,恕不一一列出了。

准的同行评议在人情和利益面前竟然如此脆弱不堪,在事涉大至动辄成百上千万甚至更多的科研项目分配,小到个人职称评审、考核这样各色各类的评价中,我们又怎么敢寄希望于同行评议能够做到自律?在同行评议仍然声名狼藉的今天,我们又怎能指望"代表作评价制度"能独善其身?

在目前的学术生态下,无论是专业性还是自律性,同行评议都很难做到。作为一种学术制度,"代表作评价制度"目前尚处于试验阶段,只是部分高校在职称评审中尝试实行,或者作为量化标准的补充①,无论教育部还是各高校都还没有制定出详备的可行性方案或规则,上述"中央两办"文件也只是规定了基本原则,而这些原则如何落到实处并大面积推广,还有待探讨。即使行政权力部门强力推行,在将"三大核心"这样的排行榜强行压制之后,1990 年代同行评议被定量评价逐出学术评价领域的那一幕还会重演,类似"三大核心"的排行榜仍会卷土重来。因为,与堕落了的同行评议相比,评价机构的期刊排行榜毕竟还有一个客观的数据在那支撑着,如果没有必要的制约手段,"代表作评价制度"在抵御"制度腐败"方面远不如排行榜靠谱。

由此可见,无论是"定量评价与定性评价相结合",还是"代表作评价制度",其主张者虽然都描绘了取代"以刊评文"的所谓定量评价的美好愿景,但要真正实现替代恐怕还有很长的路要走。在走完这条艰难的长路之前,取代类似"三大核心"这样的期刊排行榜的公正的学术共同体评价都只会是水中月、镜中花,要想获得学术界的一致认可,甚至多数认可都是不可能的。

七

行文至此,对前述观点作一小结。学术界和学术期刊界与评价机构的"缠斗"已不下二十年,之所以至今难分胜负,是因为评价机构有着自己的理论支撑和数据支持,更因为其以排行榜为主要形式的产品("三大核心")满足了行政权力的偏好,得到了行政权力事实上的采信,成功地将学术共同体的同行评议驱逐出了评价领域;但评价机构也有着自身无法克服的缺陷——因为专业能力的缺乏,其理论、数据和算法及结果都有经不起推敲之处,更因为其排行榜在学术评价实践中的运用导致了一系列严重的问题,为学术界和学术期刊界锲而不舍的抵制和讨伐源源不断地提供了动力和炮弹。在现行的科研体制之下,行政权力一方面是高于学术界和学术期刊界以及评价机构的唯一存在,另一方面却也有着为科研服务的基本属性,故既不能不顾学术界和学术期刊界的意见,不能无视排行榜在运用于评价时出现的问题,但却又无法抛弃"三大核心",因为离开了"三大核心",行政权力就难以顺畅地运行。两难之下,找到可以为学术界和学术期刊界接受的"三大核心"的替代的产品无疑是最好的选择,但这又谈何容易?在找到这样的替代产品之前,面对学术界和学术期刊界与评价机构的"缠斗",行政权力左右为难,举棋不定,从而"缠斗"也就一直持续了下来。然而,维持这样的"缠斗"是要以学术研究受到伤害为代价的。二十年来,行政权力虽不公开授权,但却不阻止"三大核心"在事涉资源分配和科研管理中发挥决定性的作用,使得学术界不得不屈从于这样的"指挥棒",继而从屈从发展到刻意迎合,某些人甚至不惜因此造假。学术不端行为的普遍发生只是学术研究遭受伤害的表现形式之一,实质性的伤害更在于学术研究的机体遭到侵蚀而不能健康发展。时至今日,这种伤害正愈演愈烈,以至于行政权力也不得不出手干预了。

2018 年权力顶层发起的清理"四唯""五唯"行动,清理范围涵盖了法律和行政法规、各类考核评价条件和指标、有关管理信息系统和工作表格等等,似可视为对多年来重"量"轻"质"的量化评价的雷霆扫穴之举,但只要细读一下这些文件就可以发现,清理行动并无意将量化评价逐出评价领

①如复旦大学 2010 年就宣布"试行'代表作'学术评价制度,2011 年扩大到文科领域,2012 年将在全校铺开",但同时申明:"复旦大学推行代表作制,并没有抛弃原有的'数论文'式的评价方式,只是把'代表作'制作为一种辅助评价手段,实行双轨评价。"(《复旦大学今年起全面推行"职务聘任代表作制"》,《新课程研究》,2012 年第 6 期)

域,甚至也不会将“三大核心”这样的排行榜完全废弃,而只是通过清理行动改变唯此独尊的地位,恢复学术共同体在评价中的应有地位,让同行评议重回评价领域并且与各种排名实现融合和互补①。可见行政权力也明白重建评价的艰难,故希望各方面共同合作,从而也为评价机构、学术界、学术期刊界在今后的评价重建中作出各自恰如其分的努力和贡献留下了充分的余地。由此,学术界与学术期刊界与评价机构的“缠斗”也许会转向合作,共建未来的评价。当然,这样的愿景是否会实现,还取决于学术生态能否改善和评价机构对自身进路的选择。

如前所述,对于行政权力的决策,评价机构要敏感得多,因为他们手中所握有的评价权力是由行政权力所赋予的,来自行政权力特别是最高权力部门对学术评价态度哪怕任何细微的改变,都足以对评价机构的前途造成决定性的影响。除了重视和迎合行政权力的意志以外,评价机构虽然极少正面回应来自学术界和学术期刊界的批评,但并不代表他们对这些批评的无视,恰恰相反,他们对这些批评的重视程度一点不亚于对行政权力的重视。他们不会不明白,来自部分学者和期刊人的批评一旦成为学术界和学术期刊界乃至舆论界的共识,是足以影响行政权力决策的。正因为如此,近年来,针对学术界和学术期刊界的批评,以“三大核心”为代表的评价机构拿出了许多应对之举,比如:加强对支撑其产品的评价理论的宣传,开拓评价数据的挖掘范围,提高评价数据内含质量,改变具体的评价方法和程序,直到让渡部分已掌握的评价权力,吸纳部分学者和期刊人参与其产品的制作,等等,以尽可能地改变自身的形象,力保自己在学术评价中的地位。这也是“三大核心”没被行政权力抛弃的重要原因之一。

另一个重要原因就是评价机构并非一无是处。多年来,评价机构实际上充当了两种角色,其一是评价数据的采集者和挖掘者;其二是利用这些数据进行运算的评价者。后一角色是对学术共同体的取代,明显越界了,因其不具备各学科专业分析能力,决定了他们不可能是合适的评价主体。但前一个角色他们还是有一定的优势的,他们二三十年来创造的最大价值不是期刊排行榜,而是他们挖掘或掌控的大量评价数据。这些数据对于学术研究、学术期刊和科研管理都是可以有独到作用的,至于“三大核心”是否发挥了这些数据的价值则是另一回事。在笔者看来,在清理“四唯”“五唯”的高压下,正是这些数据,不仅可以使制作“三大核心”的评价机构不致轰然倒塌,而且在未来学术评价机制重建的过程中还可以起到难以替代的作用。所以,对于评价机构来说,放弃评价主体,回归数据挖掘者的定位,是确保他们发挥自身作用不被驱逐出评价领域的最佳选择。但是,能否真正回归这一定位,并不是每个评价机构都愿意或都能够做到的。

由于各自的传统、拥有的资源以及既有产品形式的不同,“三大核心”的制作者在进路设计方面表现出了不同的选择。限于篇幅,本文无法对此作出全面的分析和评论,仅根据“三大核心”近年来主要的变化,对其进路选择和可能造成的结果略作评论。

在“三大核心”中,北京大学图书馆等单位联合制作的《中文核心期刊要目总览》(以下简称《总览》)问世最早,当它于1992年首次出版时,只是作为各图书馆订阅期刊的参考工具书而编制的,虽然也与评价有关,但当时的学术评价尚是同行评议的天下,其制作者大概不会想到在学术评价中会有今天这样的地位。《总览》的问世让正呼唤量化评价的学术界看到了同行评议的替代品,《总览》的功能开始越界,逐步由图书馆订阅工具书过渡到期刊评价标准再扩展成为学术评价标准。公正地说,这一功能越界行为并非《总览》制作者单一努力的结果,学术界出于对量化评价的渴望而在其中的推波助澜可谓功不可没。当然,《总览》的制作者是乐见其成的。随着《总览》评价功能的越界展开,特别是批评讨伐之声渐起以及行政权力对改革评价的回应,《总览》制作者意识

①2018年“中央两办”《关于深化项目评审、人才评价、机构评估改革的意见》指出:“SCI(科学引文索引)和核心期刊论文发表数量、论文引用榜单和影响因子排名等仅作为评价参考。……注重发挥同行评议机制在人才评价过程中的作用。”果真如此,评价机构仅靠制作排行榜独霸评价是不可能了,但作为评价的参考还是可以存在的。

到原初的遴选方法太过原始和简单了,于是有了在评价理论和制作实践方面的诸多发展。

2018 年,《总览》第 8 版(2017 年版)正式出版,使我们可以从这新版的变与不变化中一窥其进路选择。不变有二:其一,所依凭的评价理论和使用的基本方法不变。“《总览》(2017 年版)……在总体研制原则和研制方法上仍延续以往各版的做法。”①其二,以期刊排行榜为最终产品的形式不变。与不变相比,新版《总览》的变化更引人注目。首先,虽然理论不变,但制作者在对其产品作用的认定上大大地收缩了,似乎想退回到越界前的范围内:“为图书情报界、出版界等需要对期刊进行评价的用户提供参考,不具备全面评价期刊优劣的功能,不能作为衡量期刊质量的标准,更不能作为学术评价的标准。”②其次,评价指标数量有明显增加,新版《总览》较上一版指标数量增加了三分之一,达到了 16 种。再次,如果说《总览》的制作者是以搜集、整理评价数据并进行加权运算起家的,那么,随着指标数量的增加,其制作者逐渐放弃了数据的搜集工作,新版《总览》所有的数据已不再自己采集,而直接来自其他机构③。最后,参与所谓“定性评价”专家队伍的扩展④。

这些变与不变说明了什么?又将给《总览》带来怎样的命运?综合上述“四变”“两不变”来看,《总览》制作者选择的是一条以退为进的路径。对于他们说来,未见得不明白排行榜的弊病和在未来评价中必将被弱化的命运,未见得不明白数据才是安身立命的根本,但一方面要舍弃已在各界形成很大影响的《总览》这样的排行榜是难痛下决心的,另一方面,他们在此前就放弃了原始数据的采集和挖掘,现已无法弥补。所以,“两不变”是不得已的选择。在独立数据库建设基本无望、排行榜为产品唯一形式无法改变的情况下,他们的“四变”实际上是对数据采集和挖掘者角色的主动丢弃,甚至不惜在名义上放弃一贯以来所坚持的作为期刊评价乃至学术评价产品的定位,而寄希望于通过评价指标、评价方法(算法)的改变换取一个评价者的地位。应该说,这也是基于利益算计的一种可以理解的选择。毕竟科学的评价机制非短时间可以打造,能替代《总览》的产品何时才能问世,有太多的不确定因素,在此之前,凭着《总览》多年来的影响,也许还能维持相当一段时间。

晚于《总览》问世的第二大“核心”是中国社会科学院文献信息中心制作的《中国人文社会科学核心期刊要览》(以下简称《要览》),其制作原理和产品形式与《总览》基本一致,都是根据综合指标制作的期刊排行榜。不过,除了制作者单位不同以及只涉及人文社会科学期刊以外,其制作动机已明显偏向要在学术评价中分一杯羹了,因为此时《总览》已成功越界成了学术评价的一个新标准。因为形式雷同,而《总览》先行了一步,所以问世以后,《要览》的影响总不及《总览》。这一情况在 2013 年发生了重大变化,这一年,中国社会科学评价中心宣告成立,翌年即高调推出《中国人文社会科学期刊评价报告》(以下简称《评价报告》),此后,《要览》的制作虽被归入该中心名下,但一直未见更新,事实上已被《评价报告》所取代。《评价报告》虽然仍是个期刊排行榜,但该中心的出场却表现出了与以往评价机构完全不同的气势。首先,一改以往评价机构在自身评价主体认定上躲躲闪闪的底气不足,不仅俨然以天然的评价主体的身份自居,而且大有一统学术评价、号令天下的雄心:“‘以制定标准、组织评价、检查监督、保证质量’为主要职责,以制定和完善中国哲学社会科学评价标准,承担和协调中国哲学社会科学学术评价,构建和确立中国特色哲学社会科学评价体系为主要职能。”⑤其次,创造性地提出了《中国人文社会科学期刊综合评价体系(AMI)》,将评价

①张俊娥、蔡蓉华:《中文核心期刊要目总览(2017 年版)研究报告》,《中文核心期刊要目总览》(2017 年版),北京:北京大学出版社,2018 年版,第 79 - 99 页。

②张俊娥、蔡蓉华:《中文核心期刊要目总览(2017 年版)研究报告》,《中文核心期刊要目总览》(2017 年版),北京:北京大学出版社,2018 年版,第 79 - 99 页。

③详见张俊娥、蔡蓉华:《中文核心期刊要目总览(2017 年版)研究报告》,“表 4《中文核心期刊要目总览》(2017 年版)各评价指标统计源”。

④详见张俊娥、蔡蓉华:《中文核心期刊要目总览(2017 年版)研究报告》,“表 6《总览》各版参评专家数量”。

⑤《中国社会科学评价研究院简介》,http://skpj.cssn.cn/xspj/zxgk/zxjj/。

指标涉及的范围大大扩展。再次,将排行榜上期刊的等级进一步细化,创造性地在“核心期刊”之上划分出了“顶级期刊”“权威期刊”。最后,宣布将着手建立适应这一新体系的数据资源。2017年,中国社会科学评价中心更名为中国社会科学评价研究院,2018年,新版《评价报告》继续高调推出。除了承继第一版的诸多特点外,新版《评价报告》最大的改变有二:一是将其制作的期刊排行榜命名为“A刊”;二是引入专家参与评刊。

如果说《总览》制作者采取的是低调收缩的策略,那么,《评价报告》制作者采取的则是高调扩张和超越行政权力的战略。主要表现有二:其一,借助中国社会科学院在科研体制中的特殊地位,将制订与完善中国学术评价的标准和承担中国学术评价的重任加于自身。此前的评价机构虽不缺此抱负,但毕竟还希望通过行政权力授权和学术界以及学术期刊界的认可来谋求合法化,而《评价报告》制作者们则绕过行政权力和学术界,直接自我授权了。其二,其所推出的AMI综合指标体系与自我授权相配合,指标的设置已大大超出了传统核心期刊所采集或运用期刊形式数据的范围。AMI由吸引力、管理力和影响力三部分构成,除了保留了核心期刊常用指标,新增的指标主要集中于各期刊杂志社或编辑部的内部规制和运行,并辅之以扣分的手段。这不仅是对学术评价权的要求,而且染指行政管理和杂志社(编辑部)内部业务的目的已非常清晰。当然,也只有这样,他们一统评价天下的目的才有望达到。那么,他们的目的能达到吗?如果达到了又会给学术界和学术期刊界带来怎样的影响?

一个人或一个机构的雄心能否实现,还是要取决于其能力。评价研究院如果有此能力,至少要做到如下几点:第一,尽管他们可以自说自话自我授权,但这样的做法,势必打破现行主管主办制为标志的报刊管理体制,所以是绕不过行政权力和学术界以及学术期刊界的承认的,他们必须迈过从自我授权到获得正式授权和具备公信力这一关。第二,也是最根本的,要获得公信力,就必须证明《评价报告》是一个科学评价范本。尽管《评价报告》的制作者有很多创造性之举,但就其产品而言,在本质上,与《总览》和《要览》没什么区别,都是通过综合指标加权运算而得出的期刊排行榜,故而他们必须证明排行榜就是科学评价。第三,为了拓展权力边界,AMI较《要览》新增了大量的指标①,特别是涉及期刊制作业务和管理的指标,撇开其合法性不谈,在技术上也大大增加了数据采集、整理、分析的难度,仅保证数据的真实可靠,就不是一个评价机构所能轻言做到的,因此,他们必须证明在数据采集和处置方面自己具有超越常人的能力。

在笔者看来,以上三点,不仅评价研究院,任何一家以排行榜为唯一产品的评价机构都不会具此能力的。事实也是如此,老牌的评价机构如《总览》的制作者已深知满足这些条件之艰难而选择了退却;《评价报告》的制作者虽然勇气可嘉,但评价研究院成立至今,除了其所在的中国社会科学院以外,又有几家大学或科研机构认同其自我授权呢?《评价报告》问世至今,又有几家大学或科研机构将它列为评价的标准了呢?也幸亏如此,不然,仅其别出心裁推出的“顶级期刊”“权威期刊”等新的期刊级别,定会把以排行榜为形式的量化评价推向极致,定会使“以刊评文”这样扭曲了的评价歪风愈演愈烈。当然,也不能排除另外一种可能,那就是凭借中国社会科学院所具有的特殊优势,说动行政权力强行授权,使学术评价定于《评价报告》之一尊,那么,结果必定是灾难性的。除了“以刊评文”的歪风难止以外,允许不具备各学科专业研究能力的评价人员组成的评价研究院直接插手乃至指挥各期刊社(编辑部)的编辑出版业务,必将在评价机构与期刊之间建立起一种颠倒的关系,是对学术期刊和学术共同体权利的严重僭越。

其实,以“A刊”命名自己的排行榜,评价研究院并非首家。2016年4月,教育部学位与研究生

①“AMI综合评价指标体系……由3个一级指标、10个二级指标和24个三级指标构成”(中国社会科学评价院:《中国人文社会科学期刊AMI综合评价报告(2018年)》,2018年11月16日,第二篇:A刊评价报告,第1页),http://www.cssn.cn/xspj/xspj_yw/201811/W020181119593425785869.pdf。

教育发展中心决定开展全国第四轮学科评估,同时推出一项重要的改革措施,即在判断科学研究水平时,“在人文社会科学采用‘A类期刊’指标”①,以取代往届评估中使用的“C刊”(CSSCI来源期刊目录),并发布了由其编制的“A刊”目录。然而,半个月后该中心就宣布放弃“A刊”,恢复使用“C刊”。“A刊”之所以短命,不仅因为它是一个新期刊榜,更是因为它是一个即将成为学科评估标准的新期刊榜,才在遭到了远远超出该中心预料的学术界的猛烈抵制和抨击后草草收场。评价研究院以“A刊”自命期刊榜,多少有点力压“C刊”的意思,但不免有拾人牙慧之嫌。可以预料的是,一旦如评价研究院所愿,他们的“A刊”能够成为一个为行政权力所公开认可的评价标准,那么,必然引发学术界和学术期刊界与该评价机构新一轮的“缠斗”。实际上,对《评价报告》的批评之声从其问世时就有了,之所以在学术界没有出现像对“C刊”那样的激烈讨伐,只是因为在中国社会科学院以外,没几个学术单位视其为标准,也就没多少学者真的把它当回事而懒得搭理它了。

“三大核心”中,CSSCI来源期刊目录问世最晚,也是被误解误用最多的。来源期刊和核心期刊原本就不是一回事,两者最大的区别在于数据与期刊目录的关系:核心期刊制作者采集数据是为了生产期刊排行榜,排行榜是其最终也是唯一产品;CSSCI制作者评选来源期刊只是为了确定数据源,其最终产品是引文数据库,即CSSCI。所以,“CSSCI”和“CSSCI来源期刊”所指是不同的。但因为来源期刊目录也是通过遴选所得,所以也可以当作期刊排行榜来使用。CSSCI于上世纪末问世之时,核心期刊已在学术评价中大行其道,当来源期刊这个与数据相关的新的期刊目录出现之时,对引文数据库尚知之甚少的人文社会科学学者和期刊人将其视为核心期刊之一种也就不奇怪了。而作为后来者的CSSCI制作者们为了尽快在学术界和学术期刊界打出影响,也默认和乐见这种误解,于是,“三大核心”之说不胫而走。短短几年之后,CSSCI来源期刊目录即成功地反超了另两大“核心”,CSSCI来源期刊也渐渐地有了“C刊”的代称,在学术评价中的影响越来越大,而CSSCI(数据库)的影响远不如它,知“C刊”而不知CSSCI成为一种普遍现象。

CSSCI制作者很快就领教到了借“三大核心”成名的反噬之力。随着“C刊”影响日隆,大约从2010年开始,学术界和学术期刊界对评价机构讨伐的矛头所向由核心期刊转向了“C刊”,最初对“C刊”的讨伐几乎全是讨伐核心期刊的路数,令CSSCI制作者哭笑不得。痛定思痛,从此他们开始了有意识地加大作为引文数据库的CSSCI意义的宣传,并成为第一个也是迄今唯一的主动放弃学术评价主体定位的评价机构,反复强调评价只是CSSCI的附带功能,而其主要功能是为学术研究提供数据服务,从而由学术评价转向了知识服务平台的定位②。

随着近年来学术界和学术期刊界对CSSCI了解的逐步加深,CSSCI在学术研究中有了较多的运用,其对于学术研究、学术期刊的价值正在被不断地开发出来,而如何确保数据真实性和准确性等数据质量的问题也就被提了出来。与此同时,由于学术评价机制并没有改变,“以刊评文”依然如故,“C刊”在评价中的风头不仅不减,且因各界对CSSCI认同度的提高而风头更盛,“C刊”在评价中的现实作用并没有因为CSSCI制作者身份定位的转向而有改变,故而遭遇批评和讨伐仍然不可避免,如果剔除情绪性的发泄,理性批评的矛头已渐渐地集中于遴选“C刊”所依据的数据的质量瑕疵。可见,数据的质量问题已成为CSSCI制作者如何面对批评和如何实现自身价值所必须解决的迫切问题。

对于引文数据库来说,原始数据是否真实似乎是期刊和作者的问题,但在现行科研体制和评价机制之下,虚假数据实际上是和评价机构的偏好紧密关联的,造假的目的就是为了在基于排行榜的各种评价中获利,故而直接的根源还在评价机构的指标体系。在科研体制和评价机制没有根本改

①教育部学位与研究生教育发展中心:《全国第四轮学科评估邀请函》,学位中心〔2016〕42号。

②参见王文军:《检索抑或评价:CSSCI功能论析——兼论构建引文索引数据库的若干问题》,《澳门理工学报》,2013年第2期。

变的情况下,“C 刊”仍会作为评价所用的排行榜而存在,数据造假现象就很难从根源上杜绝。CSSCI 制作者要保证其引文数据库的质量,就必须尽可能减少虚假数据对数据源遴选造成的影响。具体而言,主要需解决两方面的问题:其一,虚假引用,最突出的是“互惠”引用;其二,期刊论文质量极不平衡问题。这两者解决不好都会造成来源期刊遴选的失当。

由于新版 CSSCI 来源期刊目录尚未发布,针对数据质量问题,将会有哪些遴选方法的改进尚不得而知,但从上一版发布以来 CSSCI 制作者所做的一些尝试性工作大致也可看出端倪。要防止虚假引用或互惠引用对数据质量的影响,前提是在海量数据中将它们识别出来,如今在技术上做到这一点已没有太多困难,困难的是对这些数据如何处理,是视而不见,还是剔除不实数据了事,或是把有数据不实现象的期刊排除出遴选范围?个中的分寸颇难把握。面对虚假数据问题,新版 CSSCI 来源期刊在遴选时有何新举措,我们不妨拭目而待。相对于虚假数据,鉴别和处置论文质量不一的期刊难度要大得多,这不是通过技术可以解决的,因为这已属于内容评价,只有对期刊所刊载论文进行专业分析,才能得出结论。为此,CSSCI 制作者建构了“中文学术出版物评测发布系统”,这是一个网络平台,已通过一定的途径吸纳了数以千计的各学科专业学者和学术期刊编辑加盟,参与对期刊所发表论文的抽样质量评测,评测的口径分为“创新程度、完备程度、难易程度、成果价值等四个部分”①。诚然,论文质量评价是一个十分复杂和困难的问题,该系统也许不足以彻底解决这个问题,但对于识别少数鱼龙混杂质量参差的期刊应该会有不错的效果,因为鉴别粗制滥造滥竽充数论文的难度远低于遴选最优论文,一个合格的学者和编辑都会有这个能力。该系统已于 2018 年夏上线,相信经过一段时间的摸索和数据积累,会在今后的来源期刊遴选中发挥一定的作用。

比较“三大核心”近年来的进路选择,不难发现,CSSCI 制作者之所以能主动实现身份定位的转向,是因为其所拥有并一直公开发行的引文数据库(CSSCI),这是以制作核心期刊排行榜为唯一产品的另两大“核心”所无法比拟和匹敌的。无论是《总览》制作者的低调收缩,还是《评价报告》制作者的高调突进,他们都没有改变自己作为评价主体的身份定位,其原因,除了不愿放弃手中在握的评价权力甚至还希望攫取更多的权力以外,还在于他们拿不出一个能够为学术研究和学术期刊提供数据服务的数据库,所以,他们仍会以各种手段坚持排行榜的生产和推行。可见,当年被视为同类的“三大核心”如今已分道扬镳了。

重建学术评价机制应该说是学术界、学术期刊界和行政权力多年来的共识,但直到清理“四唯”“五唯”行动发起之时,学术评价的现状却仍如笔者在十年前就指出的那样:“行政部门不便作为,学术共同体不能作为,评价机构勇于作为”②。如果说,我们可以将清理“四唯”“五唯”行动视为行政权力终于迈出了重建学术评价机制实质性的一步,那么,真正能够取代“三大核心”的必然是以学术共同体为评价主体的同行评议,而这将取决于制约同行评议的专业性和自律性条件何时具备,显然,这将是一个无比艰难的过程。可以预料的是,因为此前难以出现更好的替代品,类似“三大核心”这样的期刊排行榜的存在将会伴随这一过程的始终,只有同行评议公信力建立的那一天,它们才会真正退出学术评价。从这个意义上来说,“三大核心”的分道扬镳不失为一件好事,至少其中之一——CSSCI 的制作者已有意放弃评价主体的定位,而愿意仅作为一个评价数据的提供者而存在,由此也可以为各界提供一个认识评价机构应有角色担当的新视角。

[**责任编辑:傅　强**　ujn_fuq@163.com]

①王文军:《定量与定性相结合何以可能?——构建中国特色期刊评价指标体系的思考》,待刊稿。

②朱剑:《学术评价、学术期刊与学术国际化——对人文社会科学国际化热潮的冷思考》,《清华大学学报》(哲学社会科学版),2009 年第 5 期。