

Doi:10.11835/j.issn.1008-5831.pj.2020.08.004

欢迎按以下格式引用:马永霞,仇笏熙.“不唯”≠“不评”:论人文社会科学成果评价方式的改进[J].重庆大学学报(社会科学版),2021(3):54-66. Doi:10.11835/j.issn.1008-5831.pj.2020.08.004.



**Citation Format:** MA Yongxia, QIU Huxi. Breaking the “four-only” is not to abolish the evaluation; Improvement of evaluation method of achievements in humanities and social sciences[J]. Journal of Chongqing University(Social Science Edition),2021(3):54-66. Doi:10.11835/j.issn.1008-5831.pj.2020.08.004.

# “不唯”≠“不评”:论人文社会科学成果评价方式的改进

马永霞,仇笏熙

(北京理工大学人文与社会科学学院,北京 100081)

**摘要:**成果评价是人文社科科研评价的重要内容。破“四唯”、破“五唯”不是摒弃评价,而是要通过建立科学有效的评价方法实现“破立并举”。文章以国家社科基金年度项目的成果鉴定为研究对象,综合运用数据包络分析法(DEA)及方差分析法得出结论:以同行评议为主的定性评价在人文社科成果评价中具有一定的合理性,是人文社科成果评价应该遵循的主要方式;同时,针对阶段性成果开展计量评价可作为同行评议的重要补充,尤其是在以应用研究为主的学科成果评价中,两种评价方式的结合显得尤为重要。通过研究分析提出人文社科成果评价方式改进的新思路:将阶段性成果与最终成果共同作为评价对象;开展以同行评议为主,计量评价为辅的综合评价方式;按照研究类型不同实行分类评价;并辅之以元评价理念对现有评价体系进行动态监控与纠偏。

**关键词:**人文社会科学;科研评价;成果评价;综合评价;分类评价;阶段性成果;元评价;破“四唯”;破“五唯”

中图分类号:C12

文献标志码:A

文章编号:1008-5831(2021)03-0054-13

## 一、研究背景与文献分析

### (一)背景及意义

#### 1. 三个文件的政策指向:破“四唯”、破“五唯”

清理“四唯”“五唯”是当下乃至今后相当长一段时间的热点话题。2018年,自科技部、教育部、人力资源社会保障部、中科院和中国工程院五部门联合发布《关于开展清理“唯论文、唯职称、唯学

**基金项目:**教育部人文社会科学规划基金项目“‘双一流’背景下社会科学成果评价体系研究”(18YJA880062)

**作者简介:**马永霞,博士,北京理工大学人文与社会科学学院教授,博士研究生导师,学科带头人,Email:tmxky@tsinghua.edu.cn;仇笏熙,北京理工大学人文与社会科学学院博士研究生。

历、唯奖项”专项行动的通知》后,在科研评价及人才评价领域引发热议。人民日报曾发表评论,对清理“四唯”的内涵<sup>[1]</sup>进行阐释。学界也相继对人文社科领域清理“四唯”的着力点<sup>[2]</sup>、清理“四唯”对科研管理转型的必要性<sup>[3]</sup>等内容展开了探讨,初步形成了改进科研评价方式和评价理念的一些基础性观点。2020年2月17日,为改进科技评价体系,破除科技评价中过度看重论文数量、影响因子等指标,忽视标志性成果的质量、贡献和影响等“唯论文”的不良导向,科技部发布37号文《关于破除科技评价中“唯论文”不良导向的若干措施(试行)》,提出了九个方面共27条具体举措。九个方面主要是针对国家层面的科技管理改革举措,但同时也对承担项目单位提出了部分要求,尤其是在“突出成果质量与贡献”方面,提出不可把论文作为唯一的评价依据和考核指标,而应考虑开展通过强化分类考核导向注重标志性成果的评价模式。2020年2月18日,教育部、科技部联合发布了《关于规范高等学校SCI论文相关指标使用树立正确评价导向的若干意见》。该意见旨在破除唯分数、唯升学、唯文凭、唯论文、唯帽子的“五唯”顽瘴痼疾,破除论文“SCI至上”,从而探索建立科学的评价体系,营造良好的创新环境,加快提升教育治理体系和治理能力现代化水平。意见共十条,其中包含了建立分类评价体系、完善同行评议、规范评价活动等具体要求。

可以认为,以上三个文件所推动的改革的核心是通过建立科学合理的评价体系,营造有利于创新的学术生态环境。因此,所谓的“不唯”绝非“不评”。无论是破“四唯”还是破“五唯”,其本质都是相同的,改革重点不仅仅是在拟要破除的指标维度等具体内容上,重要的是务必要实现评价理念上的改变。通过“破唯”纠正目前仅以单一维度和指标开展评价的浮躁之风,力求将评价的重心从评价学者个人转移为评价学术成果。

## 2. 高校评价转型的实践成果寥寥

针对五部门联合发布的《通知》,以及科技部、教育部相继下发的若干“措施”和“意见”,国内部分高校通过对成果评价的反思,已逐步开展评价转型的探索与实践。以国内最早落实破“四唯”的清华大学为例。2018年已将“完善学术评价标准”列为年度重点工作,同时作为落实中办、国办关于深化“项目评审、人才评价、机构评估改革”要求的具体举措。并由校学术委员会牵头成立工作组和专家组,历时一年于2019年4月正式发布《清华大学关于完善学术评价制度的若干意见》,其中明确了学术成果在学术评价中的重要地位,鼓励教师以高质量的学术成果服务经济社会发展;并规定高质量的学术成果是“作为职务晋升和工作考评的重要学术业绩”;此外,强调要“尊重学科差异,根据各学科的特点制定相应的学术评价标准”,实现学术评价从重数量向重质量、比贡献的转变<sup>[4]</sup>。然而改革的尝试尚局限于部分高校,尤其是对于人文社科成果评价的改革探索更是凤毛麟角,绝大多数高校及科研机构仍处于观望阶段。

## 3. 科学有效成果评价体系的意義彰显

人文社科成果评价要实现“破立并举”,落实政策的过程务必处理好破与立的关系。破“四唯”“五唯”是手段而非目的,不能将原有评价体系集体废除,而是要通过厘清科研成果产生的内在逻辑,在尊重知识生产方式和学科特点的基础上,构建恰当合理的评价方式。这是由于我国科研建制以项目制为主要开展形式,国家通过经费投入支持学科发展。对科研成果开展评价是实现科研绩效测评及资源分配的依据,更是“以评促进”的重要环节。需要注意的是,人文社科领域的科研成果具有一定的隐性特质,很难像自然科学成果一样通过生产力的转化进行效益评估,但不可因为“难评价”而“弃评价”。显然,人文社科破“四唯”“五唯”的重点是建立科学有效的成果评价体系,其意

义有三:(1)成果评价是成果质量的保障。近年来,大部分高校重视各类基金项目的申报,实行目标任务制的考评方式,导致科研人员“重立项、轻结项”,使得国拨经费利用率低、优秀成果产出匮乏。因此建立以成果评价为载体的约束机制不仅能借助基金这一政策杠杆撬动优秀成果,保障科研成果质量,而且能科学地评价投入资源是否得到合理利用,实现管理目标。(2)成果评价是科研目标的导引。基金类项目的立项原则是“坚持问题导向,以严谨回答重大理论和现实问题为主攻方向”。因而对项目成果进行评价具有重要的标尺作用,一则可以判断实际产出的成果价值与预期价值是否相符;二则可遴选出高影响力的优秀成果,在同类研究中树立标杆、推举典范。(3)成果评价是对存在问题的反馈。通过成果评价,可将存在问题及专家建议反馈给研究者,有利于研究者发现研究缺陷及盲区。同时,成果评价作为项目制管理的重要环节,间接反映了基金管理流程组织与开展情况,实现及时改进与调适。因此,以破“四唯”“五唯”为契机,通过反观人文社科成果评价活动,对现有评价体系进行把脉,找出问题和症结,实现破旧立新,具有重要意义。

## (二) 学术进展

在评价理念转型的大背景下,学界对人文社科成果评价的研究之风日盛。实践中,人文社科成果的范畴界定与分类方法众多。按照成果产出的阶段,可分为阶段性成果及最终成果两类,但二者之间存在成果形式的交集,比如阶段性成果与最终成果均包含有学术论文、研究报告等成果形式。按照成果性质的类型,一般分为基础型研究和应用型研究两类。按学科层面的分类,又分为人文学科与社会学科<sup>[5]</sup>。然而其成果无论以何种分类界定,学界对人文社科成果评价的研究不外乎三种,即评价构建理论的研究(理论依据)、评价内容的研究(评什么)、评价方式的研究(怎么评)。

### 1. 关于评价理论的研究

绩效评价是人文社科成果评价的主要实现形式。早在1980年,Beamon就提出,任何企业都要通过制定适用于自身的绩效管理模式的来保证运营与协调发展<sup>[6]</sup>。随着各国科研管理的兴起,绩效管理的研究及理论成果被广泛运用于科研评价的多个方面。美国是最早开展科研绩效评价的国家,而英国政府推行的REF模式在注重绩效的前提下,更加强化了成果的质量导向。近年来,各国科研评价研究表现出以下共性:(1)以绩效评价为主要理念。包括我国在内多数国家的科研评价基本都以政府为主体或者具有政府背景,而项目制是科研开展的主要实现形式,因此对基础研究的资助及考核依赖于整体绩效评价,通过对投入产出的测算来实现管理。(2)成果评价是科研绩效评价的主要内容<sup>[7]</sup>。在美、德、英、法等科研强国的科研评价中,成果质量及创新成果的评价均在科研评价体系中占主要内容和较高权重。事实上,绩效评价是科研评价的主要实现形式,而成果评价则是科研绩效评价的主要内容。

### 2. 关于评价内容的研究

人文社科成果的形式主要体现为学术论文、学术专著、学术研究报告三类。因此,对人文社科成果评价内容的研究也主要集中于对这三种成果载体的单一评价或综合评价。目前,针对三种不同成果形式而开展的评价方法各自有别。高自龙等提出,要对人文社科学术论文开展分类评价,其方案设计中提出的按照论述体裁及一级学科进行论文分类评价的思想具有可操作性和合理性<sup>[8]</sup>;黄炜等运用绩效评价模式对2001—2013年20门人文社科学科的论文产出进行效率分析,得出论文产出效率较低的结论<sup>[9]</sup>;姚莉等针对以专著形式为成果载体的《成果文库》,提出了一套影响力评价的指标体系<sup>[10]</sup>;张玉等人构建了对中文著作学术影响力进行评价的指标体系,该体系包含4个维度

27个指标因子,并采用书评、引文、同行评议等多种方式开展评价<sup>[11]</sup>;魏庆肖采用传统计量指标与补充计量指标相结合的模式,构建了定性定量相结合的指标体系,用于完善人文社科学术专著的评价方法<sup>[12]</sup>;赖永兵通过构建模型实现对情报研究报告的综合质量测评<sup>[13]</sup>。

### 3. 关于评价方式的研究

人文社科成果的评价方式一般分为三类,即以同行评议为主的定性评价、以指标计量为主的定量评价和定性定量相结合的综合评价。其中,同行评议是国际上通用的人文社科科研成果评价的经典做法。但随着图情学和文献计量学的发展,指标计量与同行评议在人文社科成果评价中的主体地位开始处于争论及博弈状态。(1)以同行评议为主的定性评价研究。同行评议是人文社科领域认同度较高的一种评价方法,由默顿在20世纪六七十年代于《科学社会学:理论与经验调查》一书中首次提出<sup>[14]</sup>,这是同行评议的雏形和基础,之后对同行评议的研究大都基于“默顿学派”的理论进行拓展。国内较有影响力的研究是卜卫<sup>[15]</sup>等在1998年发表的《社会科学成果价值评估》中,提出人文社科的成果价值以“科学共同体”的评价作为判断的直接指标;胡明铭等<sup>[16]</sup>认为,同行评议尽管是一种主观的方法,但在人文社科评价领域尚没有更好的方法取而代之。(2)以指标计量为主的定量评价研究。人文社科成果定量评价方式主要有两种:一是根据指标因子及权重实现量化评级。比如,陆宇飞以学术著作作为研究载体,研究构建了国家社科基金后期资助项目的成果评价指标体系<sup>[17]</sup>;二是依据绩效管理理论对被评价单元进行绩效考核,其中数据包络分析法(即DEA)最为经典和实用<sup>[18]</sup>。比如,姜彤彤等将DEA法用于高校的科研绩效评估,避免了传统评价指标体系仅注重绝对数量和规模的缺陷,可用于不同评价单元的横向比较<sup>[19]</sup>。(3)定性与定量相结合的综合评价研究。定性与定量相结合的综合评价方式是近年的研究热点。2015年《莱顿宣言》提出的首条原则即为“量化评估应当支持而非取代质化的专家评审”<sup>[20]</sup>;英国科研卓越框架(REF)采用专家评议为主、计量评价为辅的综合评价方式,并且将“科研成果”作为主要考察因素<sup>[21]</sup>;郑德俊等在原有同行评议的基础上加入定量评价指标,减少了专家评议的主观影响<sup>[22]</sup>;叶继元提出以同行专家主导、量化评价辅助的方式构建学术“全评价”体系<sup>[23]</sup>;张献锋<sup>[24]</sup>探讨了通过增加评议专家群体,实现评价主体多元化,从而最大程度减少同行评议的主观性和随意性<sup>[24]</sup>。

### (三) 笔者的研究思路与期待

通过对已有文献的梳理可以发现,随着学术共同体的发展及科研环境的变化,经典的同行评议日渐暴露出一些问题和局限性,比如:随机性较强<sup>[25]</sup>、遴选过程影响评价结果的主体决策性<sup>[26]</sup>,带有偏见的“小同行”阻碍新兴研究领域的学科发展<sup>[27]</sup>等。事实上,当前的科研体制不可避免地会掺入官本位思想及小学术团体等人情关系因素,进而引发了学界对某些简单“懒政”式的同行评议的质疑及讨论。但如果单纯以指标计量来考评人文社科科研成果的价值则容易陷入“四唯”“五唯”误区。教师唯“数量论”“唯结果论”,学校“唯排名论”,导致科研成果“重量轻质”的现象日益突出<sup>[28]</sup>。

基于以上研究及其存在的问题,笔者拟选取人文社科领域具有较大覆盖面和影响力的国家社科基金年度项目(以下简称“国家社科基金项目”)的成果鉴定为研究对象。这是我国人文社科领域最高级别的国家项目,而成果评价作为其中的重要环节之一,是科学评价项目成果、提高成果质量、促进优秀成果及优秀人才产生的重要手段和保障,对人文社科整体评价具有示范作用。目前,国家社科基金项目成果评价方式是针对最终成果实行的双向匿名同行评议制度。暂未纳入阶段性成果

这一定量评价维度,缺乏对研究过程的跟踪和考核。而对国家社科基金项目的现有研究也多为单一案例分析或者局限于定性讨论的模式,缺乏大规模面板数据的实证研究。为此,笔者拟以国家社科基金项目成果鉴定数据为样本,期待研究以下问题:国家社科基金项目成果鉴定的定量评价与定性评价之间是否存在差异?差异是否具有学科特异性?阶段性成果是否有必要纳入现有成果评价体系?定量评价与定性评价之间的主次关系如何安排才能实现优缺点互补?

## 二、数据采集与研究方法

### (一) 数据采集

自2017年10月至2019年2月,研究组走访多所在京高校的人文社科科研管理部门及相关国家社科基金管理部门,调研获取了38所在京高校2013—2017年国家社科基金项目成果鉴定数据样本978份:包括免于鉴定项目74项,鉴定结果为优秀的项目45项,良好289项,合格441项,不合格88项。根据研究样本统一性原则,数据清洗去除74项免于鉴定项目,最终纳入数据分析的样本为863项。所取样本包含了国家社科基金管理划分的全部23个学科,每个项目样本数据包含阶段性成果数据及专家鉴定等级数据。样本覆盖面广,数据齐全,适于定量分析。此外,本研究收集整理了2012—2018年《国家社会科学基金年度报告》里有关年度项目成果鉴定的数据,用于分析全国整体概况。

### (二) 研究方法

本研究综合运用数据包络分析法(Data Envelopment Analysis,简称DEA法)及方差分析法等多种计量方法开展定量研究。

首先,提取整理2012—2018年《国家社会科学基金年度报告》数据,根据国家社科基金在全国范围内的成果鉴定概况进行趋势分析。

其次,以在京高校863个结项样本的阶段性成果数据为分析对象,运用DEA法对38所在京高校2013—2017年间的国家社科基金阶段性成果投入产出效率数据进行绩效分析。DEA在处理多投入、多产出的有效性评价方面具有绝对优势,其要求投入与产出指标之间具有正相关关系,即增加投入指标,不会导致产出指标的减少。从一般思维逻辑上看,国家社科基金阶段性成果投入产出数据符合DEA法的使用要求。因此,根据DEA模型选定决策单元、投入指标、产出指标,即依据国家社科基金的学科划分原则选定决策单元(DMU)为23个人文社会学科;根据数据完备性、独立性、可得性、关联性的原则,选定投入指标两项,即项目数、经费数,产出指标三项,即研究报告(以下简称研报)、专著、中外期刊论文。本研究综合运用规模收益不变的 $C^2R$ 模型和规模收益可变的 $BC^2$ 模型进行分析,评估前沿生产函数,计算各DMU的技术效率、纯技术效率、规模效率,以及规模收益情况。

再次,以863个结项样本的最终成果专家鉴定等级数据为分析对象,计算成果鉴定等级为优秀或良好的项目百分比(优良率)。将此数据与DEA结果进行比较,观察23类学科中的哪些学科存在定量评价与定性评价差异较大的现象。

最后,将以上步骤得到的定性数据及定量数据统一进行无量纲化处理,采用SPSS进行方差分析,遴选出两种评价方法处理下差异较大的学科,寻找学科分布规律。

### 三、实证分析与研究结果

#### (一) 全国项目结项与成果鉴定概况及趋势分析

选取《国家社会科学基金年度报告》(以下简称“《报告》”)作为分析的数据来源。该报告由全国哲学社会科学规划办公室统一编撰发布,是目前国家社科基金最新最全的官方统计年鉴资料。报告自2012年起发行出版,每年5月发行上一年度报告。截至2020年3月,本研究系统收集了迄今为止已发行的所有报告,选取了国家社科基金中覆盖面最广且影响力最大的年度项目作为研究对象,统计绘制2012—2018年项目结项与成果鉴定情况表(表1)。

本研究通过访谈国家社科基金项目管理人员得知,近年来基金项目的经费支持每年呈现递增趋势。该情况在表1数据中体现为:2012—2018年间申请结项数大体呈增长趋势;结项率稳定,体现了项目进度与验收情况良好。然而,优良率却出现逐年降低的趋势(图1),这表明国家社科基金尽管加大了投入支持力度,但是优良率并未随投入增多而增大。研究显示,从全国范围看,科研绩效产出情况不够理想,国家社科基金项目的研究水平及研究成果质量仍存在一定的提升空间。因此本研究认为,通过建立适当的成果评价体系以导向性提高国家社科基金项目成果质量势在必行。

表1 国家社科基金2012—2018年全国年度项目结项与成果鉴定情况

年度	申请数(项)	结项数(项)	结项率(%)	优秀(项)	良好(项)	优良率(%)
2012	1 647	1 300	78.93	194	672	66.62
2013	2 017	1 688	83.69	201	798	59.18
2014	1 713	1 430	83.48	175	638	56.85
2015	2 073	1 677	80.90	143	743	52.83
2016	2 173	1 801	82.88	142	734	48.64
2017	2 040	1 702	83.43	176	652	48.65
2018	4 587	3 861	84.17	264	1416	43.51

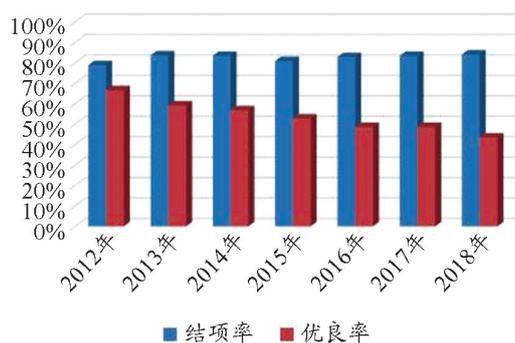


图1 国家社科基金2012—2018年全国年度项目结项率与优良率

#### (二) 定量评价与定性评价在成果评价中的差异性和同质性分析

将收集到的863个在京高校结项样本数据按照学科分类分解为23个决策单元(DMU)。首先,采用DEA法计算23个学科5年内阶段性成果的投入产出绩效值。然后,计算各学科专家评议模式下成果鉴定的优良率。最后,将两组数据(DEA数据与优良率数据)进行直观比较。在此基础上对两组数据进行无量纲化处理并计算方差,确定组间存在显著差异的学科及分布规律。

### 1. 阶段性成果的计量评价分析

表2数据是定量评价与定性评价结果的汇总。对表2中采用DEA法计算绩效的定量评价结果进行分析,得出结论如下:(1)23个学科的科研成果产出综合技术效率平均值为0.829,纯技术效率平均值为0.899,规模效率平均值0.923。反映出2013—2017年间北京地区38所样本高校的研究效率总体情况较好,项目总体上得到了良好的运行和开展。(2)4个学科技术效率有效,分别是理论经济、人口学、体育学、宗教学。这些学科的技术效率、纯技术效率及规模效率均为1,定量评价的结果说明这些单元投入合理,产出效率高,处于规模报酬最优阶段。(3)9个学科虽然技术效率无效,但纯技术效率有效( $vrsre=1$ ),包括:法学、管理学、考古学、社会学、世界历史、统计学、应用经济、哲学、中国历史。这些学科可根据其所处的规模收益阶段,适当调整投入产出规模实现DEA有效。(4)9个学科综合技术效率及纯技术效率均无效。包括:党史党建、国际问题研究、考古学、马列科社、图情、新闻与传播学、语言学、政治学、中国文学。在调研访谈中得知,这些学科多数是属于理论研究或基础性研究,成果产出周期长,阶段成果的发表存在一定困难。对于这类学科,不能简单地通过定量计数的方式来评价其科研效率,而此时同行评议无疑成为了不可替代的评价方式。

表2 北京38所高校各学科项目成果鉴定优良率及科研绩效情况(2013—2017年)

学科	项目总数(项)	优良项目数(项)	优良率(%)	crste vrste scale 规模收益			
党史党建	10	4	40	0.663	0.760	0.872	irs
法学	114	51	45	0.837	1.000	0.837	drs
管理学	64	10	16	0.938	1.000	0.938	drs
国际问题研究	28	9	32	0.737	0.785	0.939	drs
考古学	7	3	43	0.698	1.000	0.698	irs
理论经济	45	14	31	1.000	1.000	1.000	-
马列科社	20	7	35	0.587	0.594	0.988	drs
民族学	15	7	47	0.924	0.981	0.942	irs
人口学	17	5	29	1.000	1.000	1.000	-
社会学	63	17	27	0.903	1.000	0.903	drs
世界历史	7	5	71	0.859	1.000	0.859	irs
体育学	30	6	20	1.000	1.000	1.000	-
统计学	16	4	25	0.875	1.000	0.875	drs
图情	33	13	40	0.699	0.737	0.949	drs
外国文学	33	17	52	0.856	0.893	0.958	drs
新闻与传播学	42	5	12	0.718	0.754	0.953	drs
应用经济	63	17	27	0.904	1.000	0.904	drs
语言学	79	45	57	0.677	0.805	0.842	drs
哲学	66	38	58	0.952	1.000	0.952	drs
政治学	22	6	28	0.523	0.542	0.964	drs
中国历史	32	18	56	0.974	1.000	0.974	drs
中国文学	44	27	61	0.732	0.831	0.881	drs
宗教学	13	6	46	1.000	1.000	1.000	-
平均值	-	-	37	0.829	0.899	0.923	-

## 2. 针对最终成果的同行评议分析

国家社科基金最终成果评价是在综合考虑 5 位同行专家鉴定分数、等级建议、鉴定意见的基础上确定结项等级。根据结项管理规定,鉴定等级按表 3 所示划分为四类。本研究根据该等级划分标准,计算统计了 863 个样本中每个学科所包含鉴定等级为优秀或良好的项目数,并计算优良率(表 2)。从结果看出,优良率大于 50%的学科为:世界历史、外国文学、语言学、哲学、中国历史、中国文学。

## 3. 计量评价与同行评议的对比分析

将表 2 中的定量评价(技术效率)及定性评价(优良率)两组数据单独挑出进行对比分析(图 2)。单就某一观测单元(即学科)而言,两组折线图中数值点距离较近表示评价结果趋同性较强,如:外国文学、法学、图情、国际问题、哲学、考古、党史党建等学科。这类学科中定性评价与定量评价的结果变化趋势一致。而两组折线图中数值点距离较远的学科则表示评价结果差异性较强。如:体育学、管理学、人口学、理论经济等学科,这类学科中定性评价与定量评价的结果变化趋势相反。

表 3 国家社科基金成果鉴定等级量化标准

等级	量化标准(每条标准需同时具备)
优秀	①平均分在 85 分(含)以上 ②不少于 4 位专家打分在 85 分以上 ③不少于 4 位专家划等级为“优秀”
良好	①未达到“优秀”等级 ②平均分在 75 分(含)以上 ③不少于 4 位专家打分在 75 分以上 ④不少于 4 位专家划等级为“良好”以上
合格	①未达到“良好”等级 ②平均分在 65 分(含)以上 ③不少于 3 位专家打分在 65 分以上 ④不少于 3 位专家划等级为“合格”以上
不合格	低于“合格”等级的标准

总体上看,本研究得出的结果符合学界的常规认识。首先,定性和定量两种评价方式在大多数人文社科学科中的评价结果存在同质性,不存在较大的极端情况;其次,根据现有研究趋势,学界多倡导推行定性与定量相结合的综合评价模式。这与本研究得出结论一致,即同行评议在人文社科的部分学科成果评价中具有不可替代的作用,而计量评价对同行评议具有辅助作用。

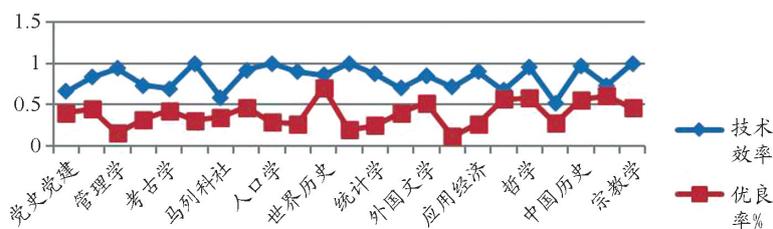


图 2 成果鉴定优良率及科研绩效对比图

## 4. 定性、定量评价方式差异较大的学科情况分析

从以上研究得知,在人文社科成果评价中,定量评价与定性评价具有同质性和差异性。为进一步探究两种评价方式差异性存在的学科分布情况,本研究利用线性函数将两种评价结果的数据进行无量纲化转换。首先按以下公式将两种评价方法得到的结果统一量纲:转换后的值=(转换前的值-样本最小值)/(样本最大值-样本最小值)。然后对两组数据进行无重复双因素方差分析。从表 4 数据结果得出:(1)组间方差值较大(大于 0.1)的学科为管理学、理论经济、人口学、社会学、体育学、统计学、应用经济。说明在这几个学科中,定性评价与定量评价的结果差异较大,阶段性成果评价在这类学科中存在的影响不可忽视。(2)组间方差值较小(小于 0.01)的学科为语言学、宗教学、新闻与传播、中国文学、世界历史、政治学、马列科社、民族学、党史党建、考古学、哲学、国际问题

研究、图情、外国文学、法学、中国历史。说明在这几个学科中,定性评价与定量评价的结果差异不大,阶段性成果评价与专家同行评议的结果具有同质性。

表4 两种评价结果无量纲化与方差分析

学科	计量评价	同行评议	求和	平均	方差
党史党建	0.293 501	0.474 576 3	0.768 077	0.384 039	0.016 394
法学	0.658 281	0.559 322	1.217 603	0.608 801	0.004 896
管理学	0.870 021	0.067 796 6	0.937 818	0.468 909	0.321 782
国际问题研究	0.448 637	0.338 983 1	0.787 62	0.393 81	0.006 012
考古学	0.366 876	0.525 423 7	0.892 3	0.446 15	0.012 569
理论经济	1	0.322 033 9	1.322 034	0.661 017	0.229 819
马列科社	0.134 172	0.389 830 5	0.524 002	0.262 001	0.032 681
民族学	0.840 671	0.593 220 3	1.433 891	0.716 946	0.030 616
人口学	1	0.288 135 6	1.288 136	0.644 068	0.253 375
社会学	0.796 646	0.254 237 3	1.050 883	0.525 441	0.147 103
世界历史	0.704 403	1	1.704 403	0.852 201	0.043 689
体育学	1	0.135 593 2	1.135 593	0.567 797	0.373 6
统计学	0.737 945	0.220 339	0.958 284	0.479 142	0.133 958
图情	0.368 973	0.474 576 3	0.843 549	0.421 775	0.005 576
外国文学	0.698 113	0.677 966 1	1.376 079	0.688 04	0.000 203
新闻与传播学	0.408 805	0	0.408 805	0.204 403	0.083 561
应用经济	0.798 742	0.254 237 3	1.052 979	0.526 49	0.148 243
语言学	0.322 851	0.762 711 9	1.085 563	0.542 782	0.096 739
哲学	0.899 371	0.779 661	1.679 032	0.839 516	0.007 165
政治学	0	0.271 186 4	0.271 186	0.135 593	0.036 771
中国历史	0.945 493	0.745 762 7	1.691 255	0.845 628	0.019 946
中国文学	0.438 155	0.830 508 5	1.268 664	0.634 332	0.076 971
宗教学	1	0.576 271 2	1.576 271	0.788 136	0.089 773

为探寻组间差异较大学科之间的共性,本研究将863个结项数据样本按照各项目研究类型的占比不同分为两类,即基础研究为主的学科和应用研究为主的学科。图3显示了全样本中23个学科的应用类/基础类占比。由图3可知,应用类研究为主(占比大于1)的学科为管理学、人口学、社会学、体育学、统计学、图情、新闻传播、应用经济;其余学科均为基础类研究为主(占比小于1)的学科。结合方差分析的结果可知,两种评价方法差异较大的情况存在于管理学、理论经济、人口学、社会学、体育学、统计学、应用经济这七类学科中,其中除了理论经济(应用基础占比小于1)外,其余学科均为以应用类研究为主的学科。说明两种评价结果差异较大的学科基本上都分布于应用类学科。因此得出结论,阶段性成果的定量评价对应用型学科存在不可忽视的作用。

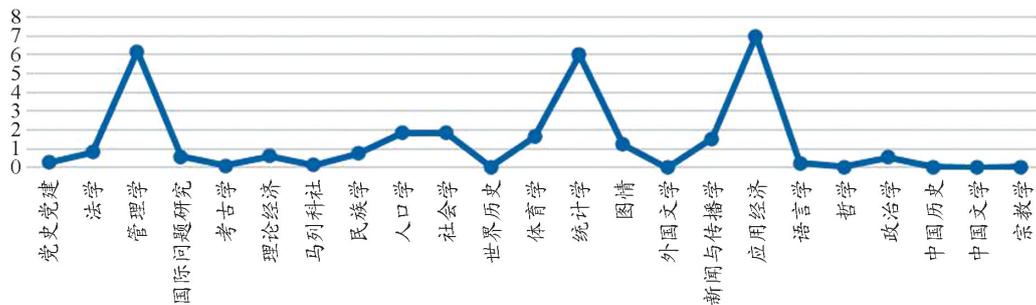


图3 各学科研究类型中应用类/基础类占比示意图

## 四、主要结论与政策建议

### (一) 主要结论

#### 1. 定性、定量评价的同质性

定性、定量两种评价方式在大部分学科评价中存在同质性。本研究对 23 个人文社科学科分别开展了定性评价及定量评价。对比后发现,两种评价方法所得结果趋同的学科共 16 个,约占 70%。说明在人文社科成果评价领域,两种评价方法对于大多数学科的评价结果差异不大,具有同质性。这也验证了,尽管同行评议存在争议,但就目前来看,这种评价方式在人文社科成果评价领域确实是无可替代的。尤其对于基础类人文社科研究,其成果具有隐形特质,成果周期较长,依靠同行专家根据以往研究经验和判断给出的评价可以最大程度地降低由单一计量评价引起的“短视效应”。因此,坚持以同行评议为主的评价方式具有合理性和科学性。

#### 2. 定性、定量评价的差异性

定性、定量两种评价方式在少部分学科评价中存在差异性。从以上研究得出结论,在 23 个样本学科中有 7 个学科,如果对其采用单一的定量或定性评价,则评价结果存在较大的差异性。具体表现为在阶段性成果定量评价中,这 7 个学科的 DEA 纯技术效率为 1,绩效测评结果表现优秀。但在同行评议时,优良率均低于平均值(37%)。因此,本研究认为,在评价以应用型研究为主的学科时,要考虑到这类学科成果产出的特点,即在研究过程中的论文发表或阶段成果的公布,具有时效性。实际上代表了较大比重的成果产出。不能搞“一刀切”,更不能为了“一碗水端平”而采用与基础研究为主的学科相同的评价方式。在这一类学科中,计量评价有必要成为同行评议的重要补充。

#### 3. 阶段性成果评价的必要性

阶段性成果的计量评价有必要作为同行评议的辅助方式纳入综合考评。本研究实证分析结果表明,在多数应用类学科中,存在同行评议部分失灵的情况(图情学、新闻传播学除外),因此,如果能在传统同行评议的基础上考虑这部分学科的阶段性成果产出状况,并将其纳入成果评价中进行综合测评,或可解决目前依靠单一同行评议对应用型学科成果评价产生的片面性。此外,对于部分基础类学科,例如理论经济,在 DEA 测评中表现出不俗的成绩(三项绩效指标均有效,位于生产前沿面上),但同行评议结果不甚理想。对于存在个别差异现象的基础性学科,假如将其阶段性成果的产出状况纳入综合考评,或可为同行评议提供数据参考。

### (二) 政策建议

#### 1. 引入元评价理念,动态监控评价偏差

“四唯”“五唯”不良倾向在以往学术评价领域长期盛行的深层次原因是对原有评价体系缺乏监督与纠偏,即缺乏对原有评价体系的元评价研究。同样,人文社科成果评价缺乏相应的元评价监控也会引起评价“偏差大”。美国州立威诺纳大学校长 R. Krugman 曾经指出,“无论哪一类评价,其结果经得起推敲的前提都必须经过元评价的检验”。目前世界各科研强国大多采用元评价的方式实现对现行评价体系的监控与纠偏。如:美国通过采用 Join Committee 提出的元评价标准对教育、科研等评价活动进行监控与再评价;日本在科研评价中运用元评价理念采取自评与他评相结合,重视戴明循环在科研成果评价中的作用<sup>[29]</sup>。我国的科研元评价尚处于理论探索阶段,尤其是针对人文社科成果评价的元评价更为缺乏。破“四唯”“五唯”不良倾向的提出,虽未提及“元评价”一词,但其实质是通过反观现有评价体系的漏洞及不足,提出修正和纠偏策略,完全属于元评价的理念范畴。如上所述,现行的人文社科成果评价多为科研绩效评价。科研绩效评价虽然能够引导科研发展的方向,但在绩效评价的过程中,由于对事物发展形态存在认知局限性,或是由于存在系统内外

环境中可变因素的影响,使得科研绩效评价出现偏差,不能达到预期的评价效果。对科研绩效评价开展元评价,能够实现科研绩效评价角色的转变,使其由评价行为的发出主体变为被评价的客体对象,对原有体系开展检验及审视。因此,务必利用破“四唯”“五唯”的良好机会,采用元评价的理念及方法开展对旧评价体系的调整,归正我国现有人文社科成果评价体系缺乏适时监控及评价反馈等弊端,从而在破旧立新中提升人文社科成果评价质量。

## 2. 引入过程性评价,妥善区分“不唯”与“不评”

国家社科基金项目的成果既包含最终成果,也包含阶段性成果。目前对国家社科基金项目成果鉴定主要针对的是最终成果,而阶段性成果仅作为数据提供给相关管理部门参考。但是阶段性成果作为反应研究过程科研绩效的重要数据,尤其是对应用类学科的项目成果鉴定具有重要的参考价值。阶段性成果包含:获奖、四大期刊转载、发表国内外论文、阶段研报、专著等指标。其中学术论文、学术专著、学术研究报告是人文社科成果形式的主要体现。首先,学术论文不仅是人文社科科研成果的主要表现形式,也是科研质量的主要评价维度。学术论文具有直观、体现研究热点、成果使用方便的优点,体现了人文社科科研的最新研究思想。“不唯论文”与“不评论文”是两个不同的概念,事实上,论文在人文社科成果中体现重要的载体作用。“不唯”的内在要求是不以论文作为单一的成果评价指标。而应注重改变评价的实质内涵,将这一指标正本清源,推行代表作制度,通过构建人文社科成果评价体系引导论文质量提升。学界对类似指标的研究也较为成熟,比如邱均平等人提出将转载作为评价指标之一的文摘评价法,是对论文水平进行间接评价的良好手段,可在很大程度上弥补由单一引文分析带来的公允性争议<sup>[30]</sup>。其次,学术专著是学术图书的主要类型,相比其他如译著、教材等图书,学术专著包含了较高的思想性及原创性,具备更强的问题导向性。对学术专著的评价是人文社科成果长效评价形式的体现。比如著名的《成果文库》多以专著形式为主。《成果文库》是全国规划办组织的对国家社科科研优秀成果进行集中推介的一项重点工作,旨在发挥优秀成果和科研人才的引领及示范作用,提升精品人文社科成果的产出效率。最后,研究报告是人文社科科研阶段性成果及最终成果较为广泛的提交形式。研究报告作为人文社科成果主要形式之一,其重要性体现在通过研究发现经济社会发展中存在的问题,并对问题的解决提出符合实际的新思路或对策,具有很大的应用价值和理论价值。对研报成果的评价主要是价值判断,即注重其适用性和可操作性。因此,通过加强人文社科的过程管理,利用好阶段性成果指标时效性强的优点,更利于知识的扩散与传播,并且数据易于计量,纳入考评的可操作性也很强。可以看出,阶段性成果评价是对科研过程的评价。目前国内评价中普遍只对结果进行考评,缺乏对科研过程的监督与管理。注重过程评价与结果评价相结合,可以更全面、更有针对性地开展科研评价,避免教师仅盯着最终成果目标来搞科研的急功近利行为。因此,可通过建立相对齐备的项目阶段性成果数据库,分析研究过程中存在的问题、研究过程与研究结果之间的关系。在原有的最终成果鉴定的基础上,增加对科研过程的绩效评价模块,形成完备的人文社科科研成果评价体系。

## 3. 引入分类评价,处理好共性与个性问题

破“四唯”“五唯”,就是要避免“一把尺子量所有”的一刀切式评价。要根据评价对象的不同,选择具有适切性的评价方法。人文社科科研按照研究类型不同分为基础研究及应用研究两类。基础研究的主要目的是为了获取基本理论,并不十分关注研究成果的直接应用问题,成果常表现为一般的原理、理论或规律。而应用研究则是为了获得特定的应用成果,比如为解决某一实际问题提供新方法、新途径等,其成果形式多以论文、研报等为主。因此,人文社科的评价要根据各学科研究类型及成果产出形式的不同,充分考虑学科特点及差异,开展分类评价。可参照人大复印报刊资料对不同学科论文的评价方式,通过提取普适性较高的标准,设置为通用指标,之后再针对不同学科或

成果体裁类型分配不同指标权重的评价方法。即采用“共识性评估标准”与不同权重配比相结合的分类评价方式,开展复合型人文社科学术成果评价模式<sup>[31]</sup>。

#### 4. 建立质量导向型综合评价体系,破除科研评价中的“四唯”“五唯”不良倾向

破除旧“四唯”“五唯”后,如何避免新构建的评价体系不会迈入另一种新“四唯”“五唯”,这是需要思考的问题。根据以上实证分析,本研究认为在开展人文社科成果评价时,要遵从学科特点,开展以专家评议最终成果为主,计量评价阶段性成果为辅的综合评价方式,且在评价方式上开展二者结合,是彻底破除“唯指标论”“唯专家论”的有效手段。

首先,定性定量相结合的综合评价是当前学界最主流、最科学的评价方式。同行评议既保证了科研质量,又对人文社科成果的隐性特质进行了最大程度的发掘,是人文社科成果评价中应当坚持的主体评价方式。而计量评价的指标具有易于量化、数据客观等特点,能规避一部分由同行评议带来的个人主观因素的干扰,对同行评议具有重要的补充,是同行评议的重要辅助方式。国家社科基金项目的成果鉴定方式虽然从表面上看是单一的定性评价模式,但实质上包含了定量评价的理念和思维,专家评分归根到底也是一种量化的评价方式。可见,定性评价与定量评价可以相互渗透、互为补充。

其次,从本文的数据分析中可知,阶段性成果的计量评价对国家社科项目成果鉴定,尤其是对于应用类学科的成果鉴定具有重要的参考价值。阶段性成果指标中诸如转载、获奖等指标,也反映了同行专家经过评价后的认可程度。因此,建立定性为主定量为辅的综合评价方式,对“阶段成果+代表成果”开展全面评价,既能克服同行评议的主观性,又能消除“四唯”“五唯”可能带来的不良影响,是实现破立并举的重要途径。通过建立人文社科成果评价新体系,确定良好科研导向,建立科研良性循环,可以使更多优秀的人文社科成果脱颖而出。

#### 参考文献:

- [1] 人才评价、学科评价要反对“四唯”[N]. 人民日报评论,2018-11-01.
- [2] 范军. 比“四唯”危害更大的是“唯项目”[J]. 社会科学动态,2018(12):15-17.
- [3] 陈先哲. 清理“四唯”推动科研管理转型[N]. 中国教育报,2018-10-26(02).
- [4] 清华大学关于完善学术评价制度的若干意见[S]. 清委发[2019]11号,2019-04.
- [5] 刘大椿. 人文社会科学研究成果评价体系研究[M]. 北京:经济科学出版社,2009:61-70.
- [6] BEAMON B M. Supply chain design and analysis: Models and methods[J]. International Journal of Production Economics, 1988, 55(3):281-294.
- [7] 马恩斯特,周静,王峻岭. 比较研究下高校科技评价指标的优化[J]. 中国高校科技,2019(10):47-50.
- [8] 高自龙,刘峰. 人文社科学术论文评价中分类方案的优化探析[J]. 江淮论坛,2011(6):8-12.
- [9] 黄炜,程慧平. 我国人文社会科学学科学术论文产出的效率研究[J]. 情报杂志,2016(4):137-140.
- [10] 姚莉,陈祖琴. 《国家哲学社会科学成果文库》影响力评价体系研究[J]. 西南民族大学学报(人文社科版),2016(6):232-236.
- [11] 张玉,潘云涛,袁军鹏,等. 论多维视角下中文科技图书学术影响力评价体系的构建[J]. 图书情报工作,2015(7):69-76.
- [12] 魏庆肖. 人文社科学术专著影响力评价指标体系构建及实现机制研究[D]. 大连:大连理工大学,2018.
- [13] 赖永兵. 情报研究报告质量评价体系的研究[J]. 情报科学,1991(1):9-18.
- [14] R. K. 默顿. 科学社会学:理论与经验研究[M]. 鲁旭东,等译. 北京:商务印书馆,2009.
- [15] 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.
- [16] 胡明铭,黄菊芳. 同行评议研究综述[J]. 中国科学基金,2005(4):251-253.
- [17] 陆宇飞. 国家哲学社会科学基金后期资助项目评价体系分析[J]. 西南民族大学学报(人文社科版),2016(6):237-240.
- [18] CHARNES A, COOPER W W, RHODES E. Measuring the efficiency of decision making units[J]. European Journal of Operational Research, 1978, 2(6):429-444.
- [19] 姜彤彤,吴修国. 教育部直属高校产学研协同创新全要素生产率研究[J]. 科技管理研究,2016(17):97-101.

- [20] 罗纳德·鲁索,全薇. 期刊影响因子,旧金山宣言和莱顿宣言:评论和意见[J]. 图书情报知识,2016(1):4-14.
- [21] Research Excellence Framework[EB/OL]. [2020-02-15]. <https://www.ref.ac.uk/2014/>.
- [22] 郑德俊,高风华. 高校人文社会科学科研绩效评价指标体系构建[J]. 科技进步与对策,2009(7):150-153.
- [23] 叶继元. 学术“全评价”分析框架与创新质量评价的难点及其对策[J]. 河南大学学报(社会科学版),2016(5):151-156.
- [24] 张献锋. 中国社科类学术期刊评价体系的若干思考[J]. 重庆大学学报(社会科学版),2015(5):116-120.
- [25] 蔡蓉华. 同行评议的难点[J]. 甘肃社会科学,2015(4):25-27.
- [26] SUN Y H, MA J, FAN Z P, et al. A group decision support approach to evaluate experts for R&D project selection[J]. IEEE Transactions on Engineering Management, 2008, 55(1): 158-170.
- [27] KENNEDY D. Disclosure and disinterest[J]. Science, 2004, 303(5654): 15.
- [28] 马永霞,仇笏熙. “双一流”建设背景下我国高校学术论文评价改革的思考[C]//2017年中国高等教育学会高等教育专业委员会第十七届年会会议论文集,2017.
- [29] 日本参议院. 政策評価法[EB/OL]. (2016-06-16) [2019-12-08]. <http://law.e-gov.go.jp/data/H13/H13HO086.html>.
- [30] 邱均平,王菲菲. 社会科学研究成果综合评价方法研究[J]. 重庆大学学报(社会科学版),2010(1):110-114.
- [31] 钱蓉. 基于同行评议的复合型人文社科学术评价:以复印报刊资料为例[J]. 河南大学学报(社会科学版),2016(5):138-144.

## Breaking the “four-only” is not to abolish the evaluation: Improvement of evaluation method of achievements in humanities and social sciences

MA Yongxia, QIU Jiayi

(School of Humanities and Social Sciences, Beijing Institute of Technology, Beijing 100081, P. R. China)

**Abstract:** Achievement evaluation is an important part in the scientific research evaluation of humanities and social sciences. Breaking the “four-only” and the “five-only” is not to abolish the evaluation, but to establish scientific and effective evaluation methods. This paper takes the achievement result of the annual project of the National Social Science Fund of China as the research object. The DEA method and variance analysis method are used to draw the conclusion as follow: qualitative evaluation mainly based on peer review has certain rationality and should be the main method in the evaluation of achievements in humanities and social sciences; quantitative evaluation of stage achievements can be used as an important supplement to peer review, especially in the applied research based subject achievement evaluation. This paper provides new ideas to improve the evaluation of achievements in humanities and social sciences: take stage achievement and final achievement together as the evaluation objects; develop a comprehensive evaluation method which mainly based on peer review and supplemented by quantitative evaluation; carry out classification evaluation according to different research types; use meta-evaluation to monitor the existing evaluation system.

**Key words:** humanities and social sciences; scientific research evaluation; achievement evaluation; comprehensive evaluation; classified evaluation; stage achievements; meta-evaluation; break the “four-only”; break the “five-only”

(责任编辑 彭建国)