

高校教师自身是否摒弃了“五唯”评价： 基于高水平研究型大学职称晋升 同行专家评议的文本分析

邵留国^{1,3}, 孔祥旭¹, 黎蕙璐¹, 张 华¹, 张仕璟^{2,3}

(1. 中南大学商学院, 湖南 长沙 410083; 2. 中南大学人事处, 湖南 长沙 410083;
3. 教育部教育立法研究基地, 湖南 长沙 410083)

摘要:在“破五唯”的改革思路中,同行专家评议被寄予了“学术守门人”的厚望。高校教师不仅是被评者,更是评审者,其评价实践是否摆脱了“五唯”路径依赖是改革成败的关键。学界对此缺乏量化实证研究。基于某985高校2020—2024年教师职称晋升函评文本材料,利用BERT模型进行多标签分类,探究“破五唯”背景下高校教师评价的整体特征演变趋势。研究表明,高水平研究型大学的评审实践呈现“科研主导、多维渐进”的改革特征,尚未实现评价体系的结构性变革,“破五唯”改革仍需进一步深化和推进。构建基于评审文本的大数据分析模型,更能捕捉评审专家的真实立场,为动态监测“破五唯”改革成效提供方法论参考。研究发现,设计优化职称评审制度、推动高等教育评价范式转型具有重要启示。

关键词:高校教师评价;“破五唯”;同行评议;文本分析

中图分类号:G521 文献标识码:A 文章编号:1005-0566(2025)09-0001-11

Have university faculty themselves abolished the “Five-Only” evaluation criteria: textual analysis based on peer expert review in faculty promotion at leading research universities

SHAO Liuguo^{1,3}, KONG Xiangxu¹, LI Huijun¹, ZHANG Hua¹, ZHANG Shijing^{2,3}

(1. Business School, Central South University, Changsha 410083, China;
2. Department of Human Resources, Central South University, Changsha 410083, China;
3. Educational Legislation Research Base, Ministry of Education, Changsha 410083, China)

Abstract: In the reform framework of “Breaking the Five-Only”, peer review is highly expected to act as the “academic gatekeeper”. University faculty members are not only evaluatees but also evaluators; whether their evaluation practices can transcend the entrenched path dependence on the “Five-Only” paradigm is pivotal to the reform’s success. However, the academic community currently lacks quantitative empirical research on this issue. Based on textual materials from the letter review for faculty professional title promotion (2020–2024) at a 985 Project university, this study applied the BERT model for multi-label classification to explore the evolutionary trends of the overall

收稿日期:2025-03-29 修回日期:2025-08-26

基金项目:教育部高校思想政治理论课教师研究专项重大课题攻关项目(24JDSZKZ09);教育部教育管理与改革专项课题项目(24JGWT0010);湖南省深化新时代教育评价改革试点项目(2023)。

作者简介:邵留国(1979—),男,山东汶上人,中南大学教授、博士生导师,研究方向为大数据分析与管理决策研究等。通信作者:张仕璟。

characteristics of university faculty evaluation in the context of the “Breaking the Five-Only” reform. The results indicate that the evaluation practices in leading research universities exhibit a reform trajectory characterized by “research-led, multidimensional, and gradual progress”. Nevertheless, a structural transformation of the evaluation system has not yet materialized, which suggests that the “Breaking the Five-Only” reform still requires further deepening and advancement. This paper develops a big data analysis model based on review texts, which can better capture the authentic perspectives of review experts. It provides a methodological reference for dynamically monitoring the reform’s effectiveness and offers important implications for optimizing the design of the professional title review system and promoting the transformation of the evaluation paradigm in higher education.

Key words: faculty evaluation; “Breaking the Five-Only”; peer review; text analysis

国家把“破五唯”作为教育评价改革的关键任务和重中之重。习近平总书记在 2018 年两院院士大会指出“人才评价制度不合理,唯论文、唯职称、唯学历的现象仍然严重”。中共中央、国务院于 2020 年印发《深化新时代教育评价改革总体方案》要求“扭转不科学的教育评价导向,坚决克服唯分数、唯升学、唯文凭、唯论文、唯帽子的顽瘴痼疾”。

在高校教师评价中,高校教师既是被评人,更是评审专家,是“破五唯”的关键主体。高校教师评价中普遍实施同行专家评议制度^[1-3],作为同行专家的高校教师在此环节拥有充分的自由裁量权。由于“五唯”评价模式易于量化的特点,一些高校教师不知不觉形成了“五唯”依赖。虽然大家不赞同“五唯”评价,但在作为评审专家时,却往往因为“五唯”评价更省事、更易形成共识而自然地使用这套评价模式。

作为“破五唯”的前沿阵地,高校在职称评审、高层次人才招聘、研究生导师遴选中进行了探索性改革^[4-5]。准确把握评审专家当前的评价取向是继续推进高校教师评价改革的必要基础工作。经过近年的改革推进,评审专家的评价是否摆脱了路径依赖,是否真正贯彻、做到了“破五唯”,学界对此问题缺乏量化实证研究。为此,本文以某 985 高校(简称 A 高校)在 2020—2024 年教师晋升中的专家评审文本为样本,分析评审专家在评价实践中的真实评价维度。这是一个理想的样本,该晋升评审并未预先规定评审专家的评审维度,评审专家可以按照自己的观点自由评审,评审文本是评审专家个人评审维度的真实表达。此外,相较问卷调查等研究方式,文本分析方法可以避免被调查人的虚假回答。本文采用词频分析、BERT 模

型等文本分析方法,测度“破五唯”后,高校教师评价中评审专家的关注点是否发生了变化。

一、文献综述

“破五唯”方案出台之后,学者从定性和定量两个角度解读了“破五唯”的内涵。从定性角度,李晓轩等^[6]认为“破五唯”并非不再看这些指标,而是不能简单地仅以这些指标加和计算,“破五唯”的核心是避免简单、机械的量化评价。王鉴等^[7]提出破除的核心在于改进结果评价,强化过程评价。从定量角度,胡业飞等^[8]基于实证得出“破五唯”政策必须遵循科学发展规律。汤建民等^[9]结合文献计量和内容分析法得出我国正处于“破五唯”的初级阶段。

教师能力的准确评价离不开合理的评价指标。Mingers 等^[10]验证了基于文献数量和影响因子的科学计量学在衡量和评估教师科研绩效方面的有效性。过度重视数量指标的趋势日益增长,建立以研究成果质量为主要基础的评价体系迫在眉睫^[11]。程德华^[12]提出将师德考核和个性化考核纳入考评体系;唐武生等^[13]认为应当从学术论文、科研项目、知识产权和成果转化等多元化成果评价高校教师的科研业绩。Xu 等^[14]提出引入国际同行评审可以改善高校教师评价中过于依赖定量指标的问题。大量分析结果显示,引导教师潜心教学、研究和创新,促进教师发展、优化师资队伍是教师评价的最终目的^[15]。易梦春等^[16]把学生对课程教学以及教师的满意度引入教师评价体系。

《深化新时代教育评价改革总体方案》(2020)以及教育部、科技部等于 2016 年、2018 年、2020 年颁布的系列文件提出实施分类分层评价。此后相关研究开始着力于从高校类别或教师类别细分重

构教师评价体系。高一琼等^[17]基于民族高校价值理念构建了民族高校教师发展性评价全要素指标体系。姚宇华^[18]研究了地方理工科大学在教师评价指标体系的研制上,提出须对标办学定位,坚持分类发展,构建科学适用的教师评价指标体系。王彬等^[19]提出高校应围绕为国育才使命、坚持五大职能目标、走好内涵发展道路构建评价体系。张玉启等^[20]构建了人才培养导向的高校教师评价体系。杨华春^[21]基于问卷调查和层次分析法,构建了服务人才引进、人事管理和人才开发的全周期高校教师评价指标体系。杨琪琪等^[22]提出需明确可测评的显性指标和不可测评的隐形指标。孔德颖等^[23]基于专家打分的关联度进行数据分析,构建了应用技术大学的教师评价体系。胡瑞等^[24]以新公共服务理论为依据,从治理目标、治理主体和治理内容入手,构建了高校教师评价改革分析框架。

综上,当前研究重点讨论了“破五唯”的内涵、“五唯”的成因和教师评价体系的完善。完善的教师评价体系只是对高校教师评价的第一步,评审专家在评审过程中是否遵循“破五唯”要求、合理地参考教师评价体系是未知的,是现有文献尚未涉及的,而这是教师评价改革过程中非常重要的环节。

二、高校教师评价文本处理

(一)数据描述

本文以 A 高校教师职称晋升函评材料为样本。晋升函评材料包括申请人提交的申请文本“申请人任现职以来研究工作概述”及附件材料、校外评审专家撰写的评审文本。每份申请材料由 3 位或 5 位校内外同行专家进行评审。申请文本以及评审文本无预先格式要求,能充分客观地体现个人的关注点,申请文本反映了高校教师认为自身什么业绩可以帮助自己晋升,评审文本反映了评审专家视角下的晋升要求。晋升函评材料如表 1 所示。

政策内化到个人实践需要一定时间,本文以《深化新时代教育评价改革总体方案》印发的 2020 年为节点,选取 2020 年上半年至 2024 年上半年的数据。该校教师晋升函评每学期开展一

次,但 2023 年只开展了下半年一次函评,共 8 次晋升函评的自评和专家评议数据,包括 1 624 位申请人填报的申请文本,5 045 位评审专家独立撰写的评审文本,覆盖理学、工学、医学、人文科学和社会科学等 5 大学科门类的 71 个一级学科。由于评价维度的显著差异,不包含临床医疗类的晋升函评材料。

表 1 A 高校教师晋升函评材料示例

申请编号	申请人工作概述	评审专家编号	评审意见
03A71...	1. 渣浆泵的设计理论...,指导硕士生毕业 2 人,在读 7 人...	1	申请人重点围绕渣浆泵...
		2	教学工作方面,申请人...
		3	由申请人的申报材料看...
072C0...	申请人任现职以来主要从事物联网和...,本科生指导方面的工作:a. 指导本科生...	1	申请人针对 RFID...
		2	申请人长期从事计算机...
		3	申请人在物联网以及边缘计算...
083E5...	任现职以来,主要开展了以下几个...,计算机软件著作权 6 项...	1	申请人主要从事飞行器法...
		2	从教学工作量来看...
		3	申请人开展的研究工...

A 高校是一所专注教育教学、科学研究、社会服务等职能的高水平综合性大学,入选 985 工程、双一流 A 类建设高校。虽然晋升函评材料仅来自于一家高校,但其评审专家来自于 54 所高校,相应地也覆盖了理学、工学、医学、人文科学和社会科学等五大学科门类 71 个一级学科下 193 个二级学科,有足够的学科代表性。当然,96.3% 的评审专家来自于 985、211 或双一流学科高校,本文研究结论更适用于高水平研究型大学。

(二)“破五唯”导向下高校教师综合评价指标体系构建

由于评价文本具有很强的内容特性,为更好地针对该领域对文本数据进行信息挖掘,参考国内外相关研究^[25-29],并结合专家咨询以及关键词聚类分析结果,本文构建了包含思政师德、教育教学、科学研究、社会服务 4 个大类共 30 个子评价指标的教师综合评价指标体系(见表 2)。表 2 中,将三级指标编为“标签编号”,标签第 1 位表示思政师德、教育教学、科学研究、社会服务 4 个主要维

度,标签后 2 位表示该细分维度的序号。

表 2 教师晋升综合评价指标体系表

一级指标	二级指标	三级指标	标签编号
思政师德	思政师德表现	思政师德表现	101
教育教学	教学工作量	课时量	201
	课堂教学质量	教学质量	202
		主持教学项目	203
	教学成果	教学成果奖	204
		教改论文	205
		出版教材	206
		教学类个人荣誉	207
		课程建设	208
		人才培养育人成效	209
	人才培养育人成效	指导学生数量	210
		指导学生获奖	211
		优博优硕	212
		发表论文	213
教学基地负责人	教学基地负责人	213	
科学研究	项目	科研项目	301
		人才计划	302
	科研论文	中文论文数量	303
		英文论文数量	304
		科研论文等级	305
		科研论文影响因子	306
		科研论文内容	307
		科研获奖	国家三大奖
	专利与成果转化	未转化的专利与成果	310
		转化的专利与成果	311
	专著	获出版基金资助	312
	社会服务	资政建言	要报、优秀网络成果
对口支援		抗疫、援藏、援非	402
横向项目		横向项目	403
学术兼职		学术兼职	404

(三)教师评价文本多标签分类步骤

具体过程以及 Bert 模型结构如图 1 所示。

1. 数据集预处理

首先进行数据清洗、去重等预处理。其次,由于一个完整的评价文本包含的评价信息内容过于复杂,本文对文本数据进行分句处理,以逗号、句号、问号、感叹号和分号作为分句的标准,将整体的文本信息拆分为不同句子的信息合集。

2. 人工标注的微调训练集构建

尽管对申请人和评审专家的文本内容不设任何要求,使其能够更真实地反映专家的态度,但也导致文本信息具有较强的内容特性,显著增加了通过传统范式提取信息的难度。为此,本文对语义分类模型进行微调,随机选择部分样本数据,根据前述综合评价指标体系对数据进行人工标注分类,得到“语句—标签”数据集。每句文本可能包含多种信息,如句子“申请人在教学和科研工作方面成绩较突出”包含教育教学、科学研究两方面评价,因此本文的标签分类为多标签分类,每个句子可以符合多个标签,也可以不符合任何标签。

为了保证人工标签标注的准确性和评价体系的科学性,训练集的每条文本皆由两名学者进行标注,并对比两者标注的结果是否具有 consistency。对存在分歧的标注样本,由第三位学者进行复核后确定最终标签。同时,随机抽取选取 10% 数据集结果交由相关领域专家进行交叉验证,结果符合数据的一致性标准,确保人工标注数据集的标签分类结果具有客观性和科学性。

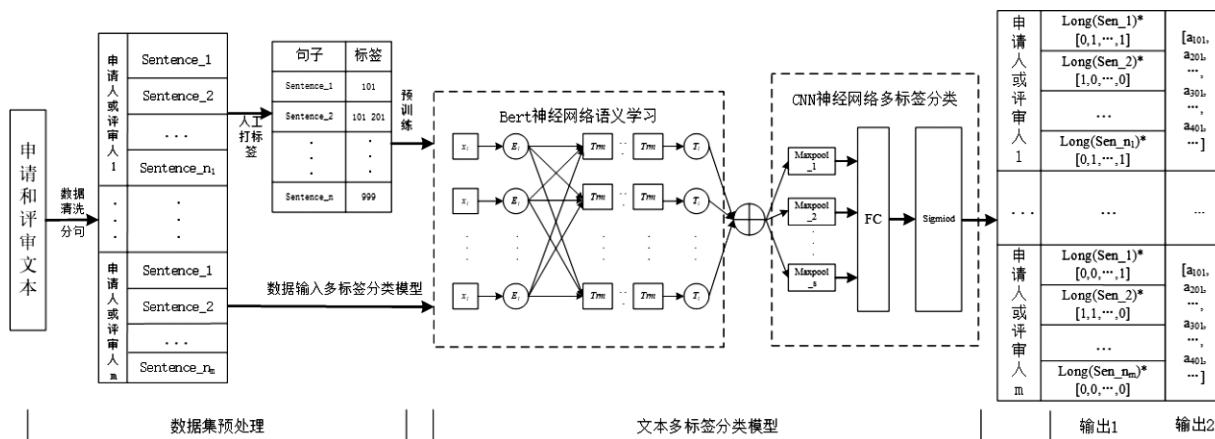


图 1 评价文本多标签分类过程结构

3. 基于 Bert 神经网络的多标签分类模型训练与分类

Bert 神经网络是一种基于语义理解的多层双向 Transformer 结构的预训练语言模型,能够捕捉文本的双向上下文信息,在各种自然语言处理任务中表现出色。本文利用前述“语句—标签”预训练数据集进行模型微调和模型检验,通过交叉验证调整参数改善模型的分类精度,得到效果较好的 Bert 多标签文本分类模型。模型由 5 层结构组成:输入层、Bert 编码层、特征提取层、多标签分类层、输出层。

4. 评价文本多标签分类结果的字符数测度

输入经过分句处理后的评价文本,利用训练好的 Bert 文本多标签分类模型判断其内容符合哪些标签,并输出其标签信息,得到如图 1 所示的结果“输出 1”。为了对该标签的信息进行挖掘,本文使用句子的字符数来表达其信息含量。根据“输出 1”的标签结果,计算每句的字符数信息。然后用该句子的字符数与其分类的结果向量 $[a_{101}, a_{201}, \dots, a_{401}, \dots]$ (其中, $a_{num} = 0$ 或 1) 相乘,得到每个句子的“标签—字符数”的信息向量。根据申请编号和评审专家编号汇总,得到每位申请人申请文本以及每位评审专家评审文本的各标签的字符数信息 $[Inf_{101}, Inf_{201}, \dots, Inf_{401}, \dots]$, 即图 1 所示的结果“输出 2”。将每位申请人或评审人所有分句的“输出 2”结果进行加和,得到该申请人或评审人的整体评价文本的有效评价标签信息,即每个人每个标签的字符信息,如表 3 所示。

表 3 部分评审文本标签信息示例

申请编号	Time	101	201	...	307	...	404	999	Long
01C4F2...	2	144	47	...	702	...	31	116	87 437
0217BB...	3	53	8	...	440	...	34	29	47 382
0320BB...	2	76	0	...	106	...	12	169	9 538
03A710...	1	46	51	...	1 036	...	42	0	35 997
03C809...	4	47	0	...	1 010	...	179	0	46 686

2020 年上半年至 2024 年上半年的每半年为一个时期,用“Time”表达;表 3 中从 101 ~ 401 列是该评审文本中被归类为相应标签的字符数;列

999 表示未被分为任何一类标签的其余字符数,比如一些过渡性描述;列“Long”表示该文本的总字符数。以表 3 中申请编号为“01C4F2...”的评审文本为例,该评审人的评审文本包含标签 101 的信息量为 144 个字符,标签 307 的信息量为 702 个字符,无标签的字符量为 116 个字符。

(四) 预训练模型分类效果

采用精确率 (precision)、召回率 (recall)、F1 值、准确率 (accuracy) 共 4 个基本指标来评价模型的分类效果。最终模型在测试集上的分类效果如表 4 所示。对于复杂多标签的文本分析而言,该模型的分类结果较好。

表 4 实验结果评价

精确率	召回率	F1 值	准确率
0.867	0.743	0.781	0.715

三、高校教师综合评价维度分析

(一) 教师评价维度重要性的度量指标设计

标签的平均字符数较高一定程度上可以综合性反映该标签受到的重视度较高,但这在实际的评价中,会存在部分标签关注度较高但无需太多文字叙述,或者标签关注度较低但需要较大的篇幅才能叙述清楚的情况,仅使用平均字符数会失之偏颇。为此本文设计以下两个指标。

1. 平均有效字符数

平均有效字符数是仅考虑对该标签进行了叙述的专家的平均字符数,测度了对该标签有关注的专家在叙述该标签时的平均叙述篇幅。其计算公式为:

$$\bar{n}_i^{effective} = \frac{\sum_{j=1}^{Num_i^{effective}} n_{ij}}{Num_i^{effective}} \quad (1)$$

式(1)中, n_i 表达标签 i 的字符数; n_{ij} 为专家 j 对标签 i 进行叙述的字符数; $Num_i^{effective}$ 为对标签 i 有叙述的人数。平均有效字符数 $\bar{n}_i^{effective}$ 测度了专家对标签 i 进行叙述的字符平均数, $\bar{n}_i^{effective}$ 较高,意味着专家认为该标签更重要。

2. 叙述比例

叙述比例是对该标签进行了叙述的人数所占

有人数的比例,测定了有多大比例的专家对该标签进行了叙述。其计算公式为:

$$\gamma_i^{effective} = \frac{Num_i^{effective}}{Num^{All}} \quad (2)$$

式(2)中, $Num_i^{effective}$ 为对标签 i 有叙述的人数, Num^{All} 为整体人数。叙述比例 $\gamma_i^{effective}$ 测度了有多少比例的专家表达了对标签 i 的关注,较高的叙述比例反映了该标签被更多的专家认为是需要进行叙述的,其受关注度更广泛。

平均有效字符数和叙述比例分别从标签受关注的重要程度和广泛程度两个方面量化了标签 i 被专家关注情况。

(二)教师评价维度整体特征分析

申请文本是申请人对自身成就的综合性展示,申请人的视角也有一定的借鉴意义。为此,本节既分析了评审文本,也分析了申请文本。

1. 评审文本整体特征分析

思政师德、科学研究和教育教学方面的叙述比例分别达到 56.29%、97.98% 及 71.99%,意味着绝大部分评审专家都对此进行了评价,但社会服务方面的叙述比例仅为 19.78%。与此同时,思政师德和社会服务的平均有效字符数仅有 26 和 30,远小于科学研究和教育教学方面。

(1)大部分专家关注到了申请人的思政师德表现,但是对此维度的评价比较简略。专家认为各申请人在此维度的表现并非是突出业绩而是基础标准,该部分的内容对晋升不具有决定作用。

(2)专家的评价仍然最关注科研方面,几乎每位专家都会进行较详细的评价,而其次是教育教学方面。

(3)专家对申请人的社会服务关注度较低,评价内容也较简略。

对评审专家在综合评价体系的细分子标签方面进行分析。分析结果如下。

(1)平均有效字符数方面,最大的标签为 307“科研论文内容”,达到 171 个平均字符,且有 91.4% 的专家都进行了叙述,因此对于评审专家而言,最主要的评审内容仍是申请人的科研具体

内容。此外,占比较高的标签有 303“中文论文数量”、302“人才计划”、211“优博优硕”及 308“国家三大奖”,对这些内容有叙述的评审专家的平均叙述字符数分别为 53、40、38 及 37。叙述篇幅较长的内容大部分都是科研方面的标签,教育教学方面更多地是与学生科研能力相关,而思政师德和社会服务方面的内容则较短。因此,可以认为,对于评审专家而言,科研方面是否符合标准仍是其表达的主要内容。

(2)叙述比例方面,除 307“科研论文内容”外,比例较高的标签有 301“科研项目”、306“科研论文影响因子”、101“思政师德”及 201“课时量”,分别达到 67.3%、56.8%、56.3% 和 48.2%,由此可以认为,这几个标签代表的方向受到更多评审专家关注。一些教育教学方面和思政师德方面的标签有更高的叙述比例而更低的字符数。这表明,专家虽然会关注这两个方面,但在实际表达中仅进行概括性总结,却把更多笔墨放在了科研能力评价方面。

2. 申请文本与评审文本的整体特征比较

由于申请人对晋升的追求,相较评审文本,其自我评价更加详细,也更加全面。申请文本在思政师德、教育教学、科学研究、社会服务 4 个维度的平均有效字符数分别为 73、277、1 125 及 86。申请人在社会服务方面的叙述比例达到了 70.5%,思政师德和教育教学方面的叙述比例也有明显提高。

在评价体系的细分标签方面,整体的结果与评审专家的结果较为相似,只有少数标签存在不同的结果。在社会服务方面的 404“学术兼职”在申请人的叙述中字符数占比和叙述比例都显著提高,该标签被申请人认为是较为重要的方面但评审人关注度不高。此外,教育教学领域中评审专家更加在意 201“课时量”和 202“教学质量”,而申请人对于 209“指导学生数量”和 210“指导学生获奖”更为关注。

整体而言,申请人在所有维度上,无论是平均有效字符数还是叙述比例都比评审专家有显著提高。

(1) 申请人往往会选择全面详细地叙述自身成绩。由于申请人存在个人简历和工作成果的不同,其叙述内容的特征可能并非显著地体现其对教师评价的关注特点,而是更多地反映其自身成果优势。

(2) 评审专家的评价不仅基于申请人个人的自述,更多地是对申请人提交的所有申报材料的综合业绩评价,因此评审专家的关注情况更能反映高校教师对教师评价的观点。

(三) 教师评价维度演变趋势分析

以“破五唯”为代表的教师评价制度改革,要求不以科研为唯一标准,要结合思政师德、教育教学、社会服务等方面开展综合评价,科研不能以数量为重,要突出质量导向。因此,本节从以下两个角度进行分析:一是分析思政师德、教育教学、科学研究、社会服务4个主要指标的时间演变趋势,探讨科研外的其他指标是否更被专家关注;二是分析科研部分的细分指标,质量、数量和内容的三者的时间演变趋势,探讨专家对科研能力的评价是否更加高质量导向。此外,由于评审文本更能反映真实评价观点,本节仅针对评审文本进行研究。

1. 综合指标的演变趋势分析

思政师德、教育教学、科学研究以及社会服务4项指标的平均有效字符数演变趋势如图2所示,叙述比例的演变趋势如图3所示。

(1) 思政师德方面,平均有效字符数和叙述比例的时间变化都较小,专家对于思政师德的关注情况并没有明显变化。

(2) 教育教学方面,相关指标的平均有效字符数和叙述比例都有明显提高的趋势,有更多的评

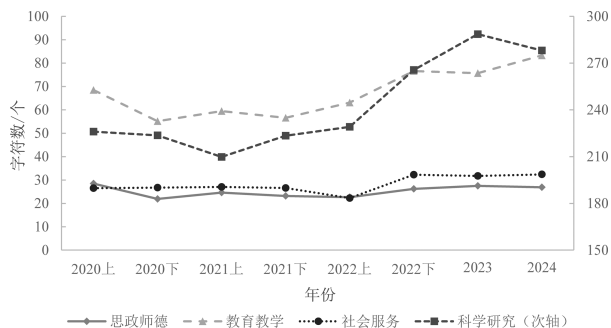


图2 主标签平均有效字符数演变趋势

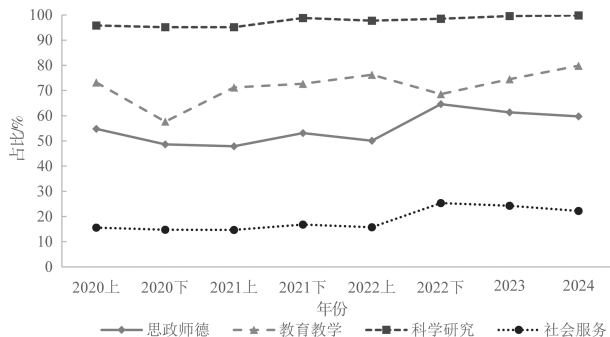


图3 主标签叙述比例演变趋势

审专家愿意对教师的教育教学问题进行更多的叙述。这在一定程度上反映了“重科研,轻教学”现象有改善的趋势。

(3) 科学研究方面,几乎所有评审专家对科研方面都有评价,其叙述比例保持稳定,同时平均有效字符数随着时间变化有所提升,说明专家对科研的重视度并未消减。

(4) 社会服务方面,平均有效字符数随时间变化相对稳定,而叙述比例在2022年后有小幅提高,有更多的评审愿意关注教师的社会服务。其中,子标签“学术兼职”对于其关注度变化的贡献较大。

2. 细分科研指标的演变趋势

更加注重科研成果的质量和内容的而非仅以数量进行评价。在评价体系中,这一导向应表现为减少对303“中文论文数量”和304“英文论文数量”的关注,更加重视307“科研论文内容”,以及305“科研论文等级”和306“科研论文影响因子”。以标签303和304反映论文数量,以标签305和306反映期刊质量,以标签307反映论文内容。图4和图5分别表示这3个指标的平均有效字符数以及叙述比例的演变趋势。

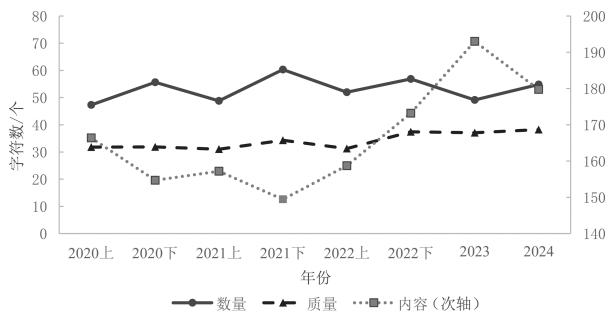


图4 科研标签平均有效字符数演变趋势

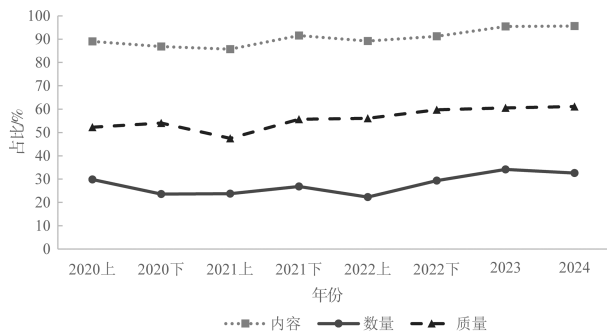


图5 科研标签叙述比例演变趋势

(1)平均有效字符数方面,论文数量和期刊质量两个指标变化趋势并不明显,而论文内容则呈现明显的上升趋势,反映出评审专家更愿意通过论文内容评价教师的演变趋势。由于论文内容所占字符数较多,这也在一定程度上解释了上节中科研内容字符数增加的现象。

(2)叙述比例方面,3个指标呈现近似平行的趋势,整体先下降后缓慢上升,反映更多的评审专家开始对论文进行多维度综合评价,但变化幅度不大。

(四)教师评价的学科异质性分析

将本文涉及的71个一级学科根据学科特点划分为理学类、工学类、医学类、人文类和社科类等5个学科类型。各时期指标的平均数如图6和图7所示。

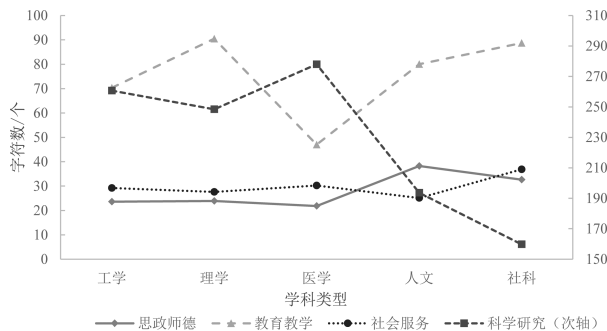


图6 各学科主标签的平均有效字符数对比

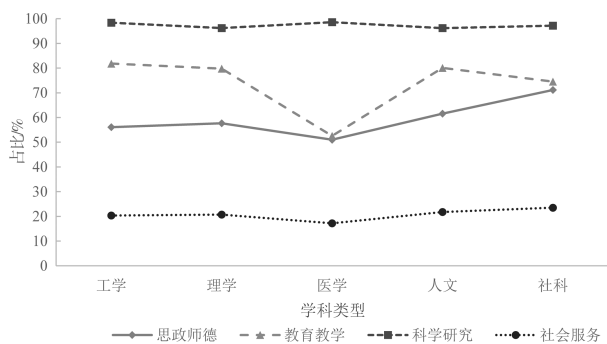


图7 各学科主标签的叙述比例对比

对比5个学科类型的主标签平均有效字符数,可以看到以下结果。

(1)思政师德和社会服务方面,5个学科类型的平均有效字符数比较接近,社科对于社会服务叙述的平均有效字符数略多于其他学科,人文学科对于思政师德叙述的平均有效字符数最多。

(2)教育教学方面,医学学科的平均有效字符数较少,理学和社科较多。

(3)科学研究方面,医学、工学和理学学科的平均有效字符数较多,而人文社科明显较少,两类学科在科研方面的评价存在较大差异。

对比5个学科类型的主标签叙述比例,可以看到以下结果。

(1)在科学研究和社会服务方面,5个学科类型的叙述比例比较接近,没有明显差异,特别是科学研究,在各类学科都非常受关注。

(2)思政师德方面,理学、工学和医学的叙述比例明显小于人文类和社科类。

(3)教育教学方面,医学由于有大量的临床教培工作,传统意义上的教学属性较其他专业更弱,因而关注度明显更低。

科学研究细分标签的学科类型对比见图8和图9。

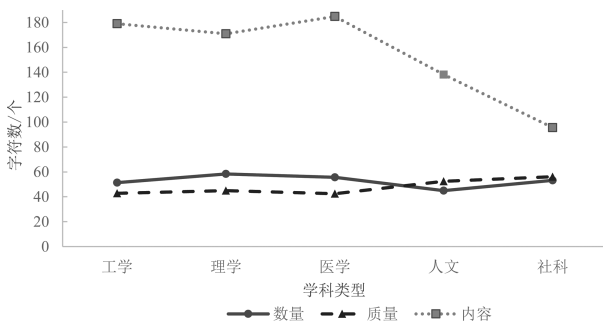


图8 各学科的科研细分标签平均有效字符数对比

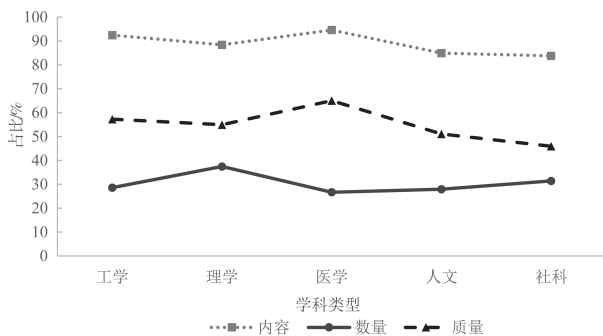


图9 各学科科研细分标签的叙述比例对比

对比各学科类型科学研究细分标签的叙述比例和平均有效字符数,可以看到以下结果。

(1)叙述比例方面,各学科类型比较接近,相对而言,人文社科在科研内容、科研质量上略低于理工医学。

(2)平均有效字符数方面,理工医学对科研内容的叙述明显高于人文社科。

综上所述,理工医类学科更看重科学研究,而人文社科类更关注教育教学和思政师德。

(五)教师评价的学校、学科等级异质性分析

将评审专家所在单位划分为“985 工程大学”“211 工程大学”和“非 985、211 工程大学”3 个等级(分别用 985、211、其他表示),计算 3 种院校主标签的平均有效字符数和叙述比例。结果显示,3 类院校无显著差异,对教师评价的关注维度具有一致性。

根据全国第四轮学科评估结果,将评审专家所在单位的学科划分为从 A+ 至 C 及 C 以下 8 个等级,分别用 A+、A、A-、B+、B、B-、C、N 表示。各等级评审专家占比为 19.1%、7.9%、23.8%、19.6%、9.0%、2.9%、3.35% 及 14.3%,评级在 B 及以上的学科,其占比较大。不同等级学科的指标如图 10、图 11 所示。

(1)在社会服务和思政师德方面,不同等级学科的平均有效字符数差异较小。

(2)在教育教学方面,平均有效字符数有较显著差异。学科等级较低的 C 和 N 学科,其平均有效字符数远大于学科等级较高的学科,而 A- 和 B+ 学科的平均有效字符数则高于 A+ 和 A 学科。

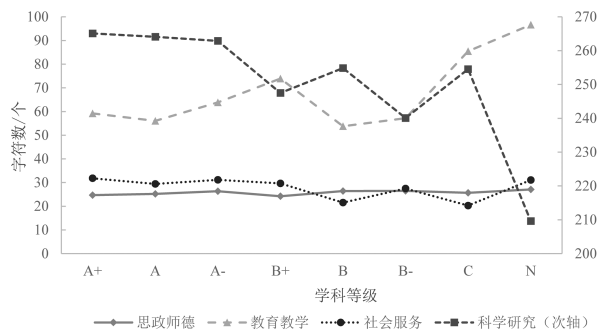


图 10 不同学科等级主标签的平均有效字符数对比

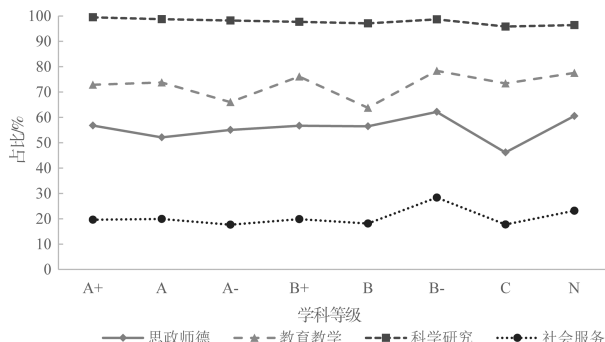


图 11 不同学科等级主标签的叙述比例对比

(3)科学研究方面,学科评级越高的学科,其平均有效字符数越高。在科学研究细分指标上,更高等级的学科对科研质量和数量的关注也更高。

(4)各个主标签的叙述比例方面,不同等级学科差异较小,而且高水平学科的整体关注度差异更小。

四、结论与政策建议

本文基于某 985 高校教师职称晋升函评材料,研判了“破五唯”政策实施的进展和效果,揭示了评审专家的评价维度是否摆脱了路径依赖、契合国家的“破五唯”要求。相较于传统问卷调查,采用文本分析更能捕捉评审专家的真实立场,本文为动态监测“破五唯”改革成效提供了方法论参考。

(一)结论

本文研究结论适用于高水平研究型大学。研究发现高水平研究型机构的评审专家“破五唯”评价取向呈现逐渐改善的趋势,在不同学科类型和学科建设层次上也有差异,但整体上看并不明显。

(1)评审专家对于思政师德的关注度不足。思政师德作为价值伦理导向的隐性素养,具有评价维度的非标定性、层级模糊性,难以通过传统指标实现“合格—良好—优秀”的结构化分层,使得评审专家难以精准区分大多数教师在思政师德方面的表现差异。

(2)评审专家对于科学研究方面的重视程度始终最高并仍呈上升趋势,科学研究仍是决定教

师晋升的最关键因素。其中,评审专家逐渐重视质量导向和实质贡献是一个积极趋势,但是这种趋势的缓慢性也暴露出“破五唯”进程的艰难。这是因为现行评价制度下,评审质量缺少被验证的环节,且评审报酬与时间投入、评审质量脱钩,评审专家不愿投入更多时间去深究论文实质贡献。此外,评审专家和评审材料专业匹配度不够高也是一个重要原因。

(3) 评审专家对教育教学和社会服务方面的关注度有一定程度的提升,“重科研、轻教学”现象在一定程度上得到改善,但在评价中的占比仍较小。主要原因是难以对二者进行高区分度评价。教育教学的最重要成果是人才,但人才培养具有显著的时间滞后性与因果隐性,不适宜较短周期的评价需求,国家级、省部级教学奖项等显性指标存在刚性配额约束,能够获奖的人极少,而针对教育教学过程的学生评教、同行评议科学性和被认可程度不高。在社会服务评价方面,评价难点与管理挑战在于难以甄别横向课题的技术含量,现有评价体系中横向课题贡献常被简化为“经费到账额”,其技术解决能力与社会效益难以测度。

(4) 评审专家的关注维度因所在学科类型不同存在差异。理工医类学科更看重科学研究,而人文社科在教育教学和思政师德的关注度有所领先。理工医类学科由于其学科特点和科研产出的相对易量化性,更看重科学研究,而人文社科由于注重思想传承、文化育人,更关注教育教学和思政师德。

(5) 评审专家的关注维度与专家所属学科的建设水平呈现出系统性关联。学科建设评级较高的学科,其评审专家更注重科学研究,相反则更重视教育教学。这一差异本质上是不同学科在既定评价制度与自身基础上所形成的差异化发展策略。科研基础相对薄弱的学科,其教师在难以产出高显示度科研成果的现实下,自然将精力更多地投入相对更有优势的教学中。

(二) 对策建议

(1) 优化政策宣导机制。加强对评审专家的政策宣讲与培训,通过规范化、结构化的评议模板和差异化的评价维度,引导专家实践。

(2) 构建思政师德分阶评价体系。将现有的“底线否决”二元模式,升级为包含“底线合规、日常增值、卓越典范”的渐进式评价体系。通过设置可测度的过程性指标,实现对教师日常师德实践的有效区分与激励。

(3) 完善评审专家遴选与履职评价机制。提升评审专家与申请材料的专业匹配度,利用国家相关机构的评审专家库和大数据、人工智能等信息技术,精准遴选“小同行”专家。优化评审效果评价和激励机制,推动评价专家投入更多精力开展以质量为导向的评价。

(4) 完善申请人提交的教育教学支撑材料。要求申请人提供教学过程数据、自我评价报告等多元化支撑材料,通过教学档案袋、增值评价等方法,推动教育教学评价从“满意度打分”转向“能力值测量”,实现对教育教学能力的精准评估。

(5) 构建横向课题三维评价体系。将横向课题评价从单一的“经费额度”标准,转变为“技术深度、经济效益、社会效益”三维标准,提升横向课题“社会服务”评价结果的被认可程度。

参考文献:

- [1] 林慧芝,刘振天. 高校教师科研评价改革的方向与重点[J]. 中国高等教育, 2021(17): 13-14.
- [2] 杨晓秋,李旭彦. 同行评议中的异常数据检测方法研究:以科研项目评审为例[J]. 中国软科学, 2016(5): 133-142.
- [3] 欧阳光华,张悦. 美国研究型大学教学同行评议的源流、论争与省思[J]. 高等教育研究, 2024, 45(5): 101-109.
- [4] 张宁娟,燕新,左晓梅,等. 构建科学的符合时代要求的教育评价制度:习近平总书记关于教育的重要论述学习研究之七[J]. 教育研究, 2022, 43(7): 4-16.
- [5] 田一聚. 我国高校教师评价改革的政策分析[J]. 江苏高教, 2022(10): 90-97.
- [6] 李晓轩,徐芳. “四唯”如何破:中国科学院研究所评价

- 的实践和启示[J]. 中国科学院院刊, 2020, 35(12): 1431-1438.
- [7]王鉴,王子君. 新时代教师评价改革:从破“五唯”到立“四有”[J]. 中国教育学刊, 2021(6): 88-94.
- [8]胡业飞,陈美欣,张怡梦. 为发展找回时间:“破五唯”政策背景下高校学科建设时间投入需求研究[J]. 国家教育行政学院学报, 2023(2): 43-51.
- [9]汤建民,凌茜. “破五唯”研究文献的分析和评价:基于文献计量和内容分析的双重视角[J]. 浙江社会科学, 2022(9): 142-150,161.
- [10]MINGERS J, LEYDESDORFF L. A review of theory and practice in scientometrics[J]. European journal of operational research, 2015, 246(1): 1-19.
- [11]CHENG J P, LI X X, XU F. Science-evaluation reform on the road in China[J]. National science review, 2018, 5(5): 605-615.
- [12]程德华. 高校教师考核评价研究评述[J]. 教育评论, 2017(9): 56-60.
- [13]唐武生,于洪莉,宋立军,等. 地方高校科研业绩评价量化体系研究[J]. 科研管理, 2016, 37(S1): 324-327.
- [14]XU F, LI X X. Practice and challenge of international peer review: a case study of research evaluation of CAS centers for excellence[J]. Journal of data and information science, 2019, 4(3): 22-34.
- [15]赵燕,汪霞. 对我国大学教师评价制度的反思与建议[J]. 高校教育管理, 2019, 13(2): 117-124.
- [16]易梦春,吴红斌. 我国高质量高等医学教育体系建设的成效评价及推进策略[J]. 高等教育研究, 2023, 44(9): 74-88.
- [17]高一琼,张男. 教育评价改革背景下民族高校教师评价体系构建[J]. 民族教育研究, 2021, 32(6): 11-17.
- [18]姚宇华. 地方理工科大学教师评价的困境表征与路径选择[J]. 中国人民大学教育学报, 2024(5): 31-43.
- [19]王彬,陈江. 破“五唯”背景下高校教师评价机制改革研究:以习近平人才观为视域[J]. 财会通讯, 2023(19): 171-176.
- [20]张玉启,兰正彦. 构建人才培养导向的高校教师评价体系:困境、动因与路径:以重庆15所本科高校专业技术职务评聘的政策分析为例[J]. 中国大学教学, 2024(6): 62-67.
- [21]杨华春,任南,吴洁. 高校教师评价指标体系的改进思考:以国际视野下的全周期为视角[J]. 中国高校科技, 2021(8): 30-35.
- [22]杨琪琪,蔡文伯. 绩效问责制导向下高校教师评价指标体系的陷阱及优化策略[J]. 复旦教育论坛, 2021, 19(4): 85-91.
- [23]孔德颖,张春明. 应用技术大学教师评价指标体系构建与应用[J]. 职业技术教育, 2022, 43(8): 70-74.
- [24]胡瑞,李彩云. 高校教师评价改革的“破”与“立”:基于新公共服务理论的分析框架[J]. 现代教育管理, 2024(8): 99-107.
- [25]谢寿安,赵长江. “破五唯”背景下高校教师评价体系构建[J]. 教育教学论坛, 2024(9): 5-8.
- [26]谢慧,李如意. 《教师法》修订视域下教师发展性评价的应然路向[J]. 当代教育理论与实践, 2023, 15(5): 137-142.
- [27]王若雅,周耀东,李宛华. 中国高校专利转化激励政策的实施效果研究[J]. 中国软科学, 2023(4): 20-32.
- [28]亢延锟,郭家宝,胡志安,等. 创新驱动、激励机制与高校科技成果转化:以省部共建国家重点实验室为例[J]. 管理世界, 2025, 41(3): 50-76.
- [29]丁大尉,李正风,罗昊雯. 科学治理视域下的我国开放科学实践:现状、动力与对策[J]. 中国软科学, 2024(1): 59-66,98.

(本文责编:默 黎)