



大学教育科学  
University Education Science  
ISSN 1672-0717, CN 43-1398/G4

## 《大学教育科学》网络首发论文

题目：中国研究生教育外部评估体系：政策检视、问题剖析与完善路径  
作者：周文辉，赵金敏  
收稿日期：2022-07-31  
网络首发日期：2022-11-07  
引用格式：周文辉，赵金敏. 中国研究生教育外部评估体系：政策检视、问题剖析与完善路径[J/OL]. 大学教育科学.  
<https://kns.cnki.net/kcms/detail/43.1398.G4.20221104.0040.006.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 中国研究生教育外部评估体系：政策检视、 问题剖析与完善路径

周文辉 赵金敏

**摘要：**中国研究生教育评估体系呈现出以外部评估为主导的本土特色，科学的外部评估体系对研究生教育内涵式发展至关重要。历经40年探索与变革，我国研究生教育外部评估体系在制度建设方面取得了一定成效，但尚存在一些现实问题有待改进，具体表现为：指标体系的认同度有待提升、数据采集增大高校工作负担、评估引发参评对象同质化、评估的诊断改进功能较难发挥。审视这些问题，其背后的原因主要是：资源质量观主导评估标准、缺乏统筹规划和归口管理、技术理性主导评估设计、忽视高校教师的主体性。面向新时代教育评价的改革需求，应培育第三方评估机构、搭建监测数据平台、强化研究生教育督导、开展元评估，以此建成更加成熟的研究生教育外部评估体系。

**关键词：**研究生教育；外部评估体系；政策检视；第三方评估；监测评估；元评估

**中图分类号：**G 643

**文献标识码：**A

**文章编号：**1672-0717(2022)06-0052-10

研究生教育是国家人才战略的重要构成。为增强本国高端人力资源的竞争力，各国政府均在不同程度上介入研究生教育评估活动，希望借助外部评估来考察研究生培养单位的内部质量。研究生教育外部评估是指“由校外全国性或区域性的研究生教育质量保障机构对培养单位所进行的质量审核、质量评估和社会评价活动”<sup>[1]</sup>。1981年实施《中华人民共和国学位条例》后，国务院成立了学位委员会和学科评议组，实行国家集中统一的学位评审制度。我国研究生教育外部评估随之诞生，由此建立了研究生教育准入机制，并着手设立学位授权点的质量检查制度和评估机构。进入21世纪之后，研究生教育事业面临着新的考验，提升教育质量成为我国研究生教育的重要目标。在保障教育质量的基础上，研究生教育外部评估体系日益重视评估的质量提升作用，开展了学位论文抽检、优秀论文评选、一级学科整体水平评估、学位授权点定期评估等项目，形成了常态化、

周期性的评估机制。近十年来，随着专业学位研究生教育蓬勃发展，专业学位评估工作愈发彰显出现实价值，面向专业学位的评估项目应运而生。研究生教育外部评估的评估对象由学术学位延伸至专业学位，评估主体由教育行政部门拓展到第三方评估机构，围绕着学位点、学科、专业、学位论文等要素，分别在研究生教育的起点、过程和终点设置了相应的评估项目，形成了对研究生教育全阶段的完整评估体系。

历经40余年建设与发展，具有中国特色的研究生教育外部评估体系初步成型。从我国研究生教育评估的现状来看，外部评估是评估体系中的核心内容。与其他国家相比，我国研究生教育外部评估主要由国家行政部门主导，且外部评估对内部评估具有强大的引导和约束作用是中国研究生教育评估的本土特征<sup>[2]</sup>。然而从实践效果来看，外部评估在回应质量问题的同时也产生了诸多非预期后果。如何保证研究生教育外部评估的科学性和有效性，尤其是

**收稿日期：**2022-07-31

**作者简介：**周文辉(1970—)，男，江苏泰兴人，管理学博士，北京理工大学研究生教育研究中心副主任、研究员，主要从事研究生教育、教育政策研究；赵金敏，北京理工大学人文与社会科学学院博士研究生。北京，100081。

如何达成“以评促改”的初心，是当前教育管理部门以及高校面临的艰巨挑战。鉴于此，本研究基于具有代表性和影响力的评估项目方案，呈现外部评估体系整体上的政策现状；采用半结构化访谈的方式对国家教育行政部门的评估行动者、高校管理部门负责人、院系行政人员以及高校教师进行深度访谈，以了解评估政策的实践面貌和制约评估效果的现实问题；在此基础上，思考与挖掘问题背后的成因，探讨完善评估体系的可行路径。

## 一、中国研究生教育外部评估体系的政策检视

目前，我国研究生教育已基本形成了常态监测、动态调整、多元综合的外部评估体系。其中，学位授权审核、学位授权点合格评估、学科评估、专业学位水平评估、学位论文抽检、“双一流”建设及其成效评价等评估项目，对研究生培养单位具有较强引导和约束作用，在展现研究生教育外部评估体系的现状与特征时有着一定的代表性。鉴于此，本研究通过对以上项目的评估要素进行梳理，并展开细致描述和分类归纳，希望呈现研究生教育外部评估体系的政策现状。

### （一）外部评估项目的主客体、类型与周期

目前我国研究生教育外部评估形成了相对稳定的政策方案，为评估项目提供了制度保障和实践

指导。就评估主体而言，外部评估体系形成了国家教育行政部门组织实施、专业协会评议、高校积极参与的多元协同模式，具体负责落实评估活动的官方机构或半官方机构主要包括国务院学位委员会、教育部学位与研究生教育发展中心、省级学位委员会，直接参与教育评估的人员一般来自国务院学位委员会、省级学位委员会、国务院学科评议组、全国专业学位研究生教育指导委员会、高校领导者、学科专家和其他行业专家。就评估客体而言，学位授权审核和学位授权点合格评估面向学位授权单位、研究生学位点，学科评估、专业学位水平评估分别用来评估一级学科和专业学位，学位论文抽检针对研究生学位论文，“双一流”建设遴选评价面向国内高水平大学与学科，“双一流”建设成效评价则针对已入选“双一流”建设的高校与学科。为实现质量保障与质量提升的双重目标，现有的外部评估项目涵盖了合格评估、水平评估和选优评估多种类型。学位授权审核、学位授权点合格评估和学位论文抽检类属合格评估，旨在保证研究生教育的质量底线；学科评估、专业学位水平评估和“双一流”建设及其成效评价致力于提升研究生教育的整体实力和国际竞争力。在评估周期方面，当前实施的外部评估主要是周期性的常态化评估。伴随着评估活动的更新迭代，评估方案也呈现出历时性和创新性，推动了外部评估体系的完善与发展（见表1）。

表1 研究生教育外部评估项目的评估主客体、类型与周期<sup>①</sup>

评估项目	评估主体	评估客体	评估类型	评估周期
学位授权审核	国务院学位委员会、省级学位委员会、部分高校自主审核	新增博士硕士学位授予单位、新增博士硕士学位点、自主审核单位新增博士硕士学位点	合格评估	3年
学位授权点合格评估	国务院学位委员会、省级学位委员会	学位授权满3年的新增学位授权点（专项合格评估）、授权满6年的学位授权点（周期性合格评估）	合格评估	6年
学科评估	教育部学位与研究生教育发展中心	具有博士硕士学位授予权的一级学科	水平评估	4年
专业学位水平评估	国务院教育督导委员会办公室组织实施，学位中心开展评估	2015年12月31日前获得专业学位授权、通过专项评估或合格评估的30个专业学位类别	水平评估	4年
学位论文抽检	国务院教育督导委员会办公室、省级教育行政部门	上一学年授予研究生学位的论文	合格评估	1年
“双一流”建设遴选与成效评价	教育部、财政部、国家发展改革委组织实施，教育部研究生司负责开展评价	高水平大学和学科	选优评估	5年

<sup>①</sup>评估要素概况基于《博士硕士学位授权审核办法》《学位授权点合格评估办法》《第五轮学科评估工作方案》《全国专业学位水平评估实施方案》《博士硕士学位论文抽检办法》、《统筹推进世界一流大学和一流学科建设实施办法（暂行）》和《“双一流”建设成效评价办法（试行）》等政策文件整理得到。

## (二) 外部评估项目的评估标准与实施程序

评估标准及其实施程序关乎评价结果的科学性和可靠性。我国研究生教育外部评估的标准体现在评估指标体系中,针对特定的评估内容,各评估项目设置了一级指标(见表2)。考虑到参评对象的异质性,评估项目的指标体系呈现出分类设计的趋势。例如,学科评估按一级学科设置了99套指标体系,根据各学科特色分别设置17~21个三级指标;专业学位水平评估针对专业学位类别建立了36套指标体系,按专业学位特点分别设置15~16个三级指标;学

位授权点合格评估的指标体系由参评对象结合院校与专业特色自主制定。在评估的实施过程中,外部评估不断创新评估工具,以指标体系为纽带,借助调查问卷和数据平台,将教育评估专家、各方利益相关者和评估数据结合起来。尽管不同评估项目在实施程序的细节上有所差异,但均详细规定了各级各类参与者在评估过程中的职责,建立了权责清晰、分工明确、执行有力的评估管理体系。该体系总体上包括确定评估范围、制定评估方案、收集信息、组织评议和发布评估结果几个环节(见表2)。

表2 研究生教育外部评估项目的评价标准与实施程序<sup>①</sup>

评估项目	一级指标	评估程序
学位授权审核	学位建设时长; 办学定位与特色; 师资队伍与水平; 人才培养与质量; 科学研究与贡献; 条件支撑与管理	高校向省级学位委员会提出申请—省级学位委员会核查材料并确定申请资格—省级学位委员会组织评议,提出拟推荐名单—国务院学位委员会组织专家进行复评与审批
学位授权点合格评估	师资队伍; 学科方向; 科学研究; 社会服务; 学术交流; 条件建设和制度保障; 人才培养数量质量和特色	学位授予单位制定自我评估实施方案—组织自我评估—定期采集评估信息—校内外专家开展评议—给出自评结果并形成《学位授权点自我评估总结报告》—向学位办报送学位点基本状态信息—上传评估结果与改进方案—学位办抽评学位授权点—发布评估结果
第五轮学科评估	人才培养质量; 师资队伍与资源; 科学研究(艺术/设计)水平; 社会服务与学科声誉	自愿申请—信息采集—信息核查—信息公示—反馈复核—专家评价—问卷调查—结果形成—结果发布—诊断分析
专业学位水平评估	教学质量; 学习质量; 职业发展质量	参评确认—信息采集—信息核查—专家评价—问卷调查—权重确定—结果形成与发布—持续改进
学位论文抽检	选题与综述; 创新性及论文价值; 科研能力与基础知识; 论文规范性	博士按10%、硕士按5%的比例抽取学位论文—制定学位论文评议要素—专家评议—反馈与公布结果
“双一流”建设遴选与成效评价	人才培养; 科学研究; 社会服务; 文化传承; 师资队伍建设; 国际交流合作	专家委员会论证确定认定标准—产生拟建设高校名单—教育部等部门确定最终建设名单—入选高校编制建设方案—每轮建设中期开展建设高校自我评估—每轮建设期末开展建设周期成效评价并形成综合评价结果

## (三) 外部评估项目的结果发布与应用

评估结果的发布影响着大学的社会声誉甚至大学排名,评估结果的应用则关系着外部评估项目的影响力,因此评估结果的发布和应用容易引起广泛关注。在结果发布方面,评估项目依据评估目标的差异,采取了不同的结果发布形式。一般而言,合格

类型的评估项目,其结果公开程度较高。例如,学位授权审核需面向社会公开高校的申请材料,当中包括申请单位在师资队伍、培养条件等方面的详细信息;高校在学位授权点合格评估中需要公开发布《研究生教育发展质量年度报告》和《学位授权点建设年度报告》。类属水平评估、选优评估的项目通

常以划分等级或区间的形式来呈现评估结果，不发布参评对象的详细信息。如学科评估与专业学位水平评估通过等级分档的形式发布结果，“双一流”建设成效评价以区间和梯度分布呈现综合评价结果（见表3）。在结果应用方面，外部评估已经成为我国配置研究生学位点、招生计划、拨款数额等教育资源的重要手段之一。例如，教育行政部门将学位

授权点合格评估结果作为监测“双一流”建设、地方高水平大学及学科建设项目的重要内容；“双一流”建设遴选及其成效评价与高校的资金、政策支持密切相关；学位论文抽检结果是确定学位授予资格、导师招生资格的重要依据。由此，研究生教育外部评估使高校产生质量保障的压力，促使其加强内部监管和建设。

表3 研究生教育外部评估项目的评估结果发布与应用<sup>①</sup>

评估项目	评估结果的发布	评估结果的应用
学位授权审核	向社会公开高校的申请材料和审核结果	学位点布局调整、学位点发展规划的重要依据
学位授权点合格评估	公布各培养单位《研究生教育发展质量年度报告》《学位授权点建设年度报告》以及最终评估结果	监测“双一流”建设和地方高水平大学及学科建设项目的重要内容，研究生招生计划安排、学位授权点增列的重要依据
学科评估	将前70%的学科分为9档向社会公布（前四轮学科评估）	为参评学科和高校提供诊断分析报告、为各级政府提供总体分析研究报告
专业学位水平评估	将每个专业学位类别前75%的参评单位分为9档向社会公布	排查专业学位研究生教育薄弱环节和存在问题
学位论文抽检	以“合格”“不合格”2档向学校公开	学位授权点合格评估的重要指标，对不合格的单位进行质量约谈和责令整改，本单位导师招生资格确定、研究生教育资源配置的重要依据
“双一流”建设遴选与成效评价	公布建设高校及学科的建设方案，以区间和梯度分布等形式，呈现综合评价结果，不计算总分、不发布排名	对建设高校和建设学科予以资金、政策、资源支持，成效评价结果作为下一轮建设范围动态调整的主要依据

## 二、中国研究生教育外部评估体系的问题表征

现阶段我国研究生教育外部评估体系在制度建设方面取得了一定成效，对于研究生教育的实践改进也发挥了积极作用。但相比总结成就，更重要的是发现其中的不足。基于对政策现状和访谈资料的归纳分析发现，我国研究生外部评估体系在评价指标、评估数据采集和评估结果方面尚存在一些不足，有待改进。

### （一）指标体系的认同度有待提升

在不同国家的外部评估中，评估指标体系的详细程度存在较大差异，一些国家仅为评估提供一级指标，由参评对象自主选取二级指标和三级指标。除“双一流”建设评价之外，我国多数研究生教育评估活动设计了明确的指标体系，每个指标下包含多个子指标，并赋以具体的分值和权重。在详尽的评价系统中，指标体系的科学性对于评估结果的信服力至关重要。然而，当前评估活动的指标体系设计还存在缺陷，致使指标体系的认可度未能达到预期、评估结果遭受批评。

一是过多使用教育投入代替教育产出。研究生教育评估最终指向教育质量，体现为人才培养、科学研究和社会服务等大学职能的履行状况。因此，评估需要围绕着人才培养质量、科学研究水平和社会服务能力等教育产出来建立指标体系。但我国研究生教育评估通常使用师资队伍、学术平台等教育投入指标，以及科研成果等部分可视化的教育产出指标来代表整体的教育质量，社会服务和人才培养在指标体系中的地位需要进一步凸显。

二是评估指标共线性较高。各个评价指标之间应相对独立，但我国多个研究生教育评估指标体系存在共线性明显的问题。例如，学位授权审核中“学位建设期”与“硕博学位授权时长”、“科学研究基础”与“学科基础”之间具有明显的共线性；学科评估中，师资队伍的人才计划数据与科研论文、项目和奖励等指标之间重复度较高。这会导致部分指标被重复计算，有损评估结果的真实性和全面性。

三是“重量轻质”的评价思维难以克服。伴随着数字测量在评估指标体系中的大面积、高频度使用，以数字为规训工具的量化思维主导了研究生教

育质量评价,评估工作沦为“计件”、教师的专业工作沦为“挣工分”<sup>[3]</sup>。尽管近年来研究生教育评估有意识地完善评估方法,评估体系也表现出客观评价为主、主观评价为辅,量化评价为主、声誉评价为辅的特点,但在本质上仍难以克服结果导向的量化思维。这种数字依附容易导致评估体系僵化、人的主体地位淡化,最终使评估成为“数字游戏”<sup>[4]</sup>。

## (二) 评估数据采集增大高校工作负担

在外部评估中,高校不仅是被评议者,同时又充当参与者的角色,需要配合评估主体为其提供所需的评估信息。目前我国实施了多项周期性的常态化外部评估项目,可以说研究生培养单位常年处于评估活动的监测之下。如学位论文抽检的评估周期为1年、学位授权点合格评估需要培养单位每年编制《研究生教育发展质量年度报告》和《学位授权点建设年度报告》,高校在每轮“双一流”建设中期和终期均需开展自我评价。与此同时,我国研究生教育的培养规模迅速扩大、学位类型日益丰富,加之评估程序日益复杂、评估指标更加精细(如学科评估设置了99套指标体系,按学科特色分别设置了17~21个三级指标)。这些变化使高校的评估工作量极大增加,参评单位需要收集与整理海量的评估信息,花费在评估活动上的时间成倍增长。

虽然有的评估组织者将自身定位为第三方非行政性评估人员,但他们仍隶属于政府部门或事业单位,具有半官方性质,这导致评估人员兼具“督学”和“督政”的双重角色,使教育评估长期伴有行政主导色彩<sup>[5]</sup>。在行政问责性的模式下,评估会成为高校在一段时期内的工作重点,参评对象始终处于自上而下的问责压力下,并全力迎合评估的要求。日益增加的评估任务层层贯彻落实到高校行政部门、院系、学科、个人,导致高校在一定程度上牺牲办学资源、教育教学秩序、教师的时间和精力。有学者指出:“每次评估都要求填写大量的表格,要求参评学校准备大量的佐证材料,这些繁琐的工作浪费了基层管理工作者和教师很大一部分时间。”<sup>[6]</sup>评估“进校”不仅增加高校人员的工作负担、影响其日常活动,长此以往还容易导致参评对象失去质量建设的积极性和主动性<sup>[7](P16)</sup>。

## (三) 评估引发参评对象同质化

我国研究生教育外部评估以政府为主导,评估活动天然地具备制度合法性。参评对象为了确立自身的

合法性,需要接受和采纳评估项目提出的要求<sup>[8]</sup>。由于我国研究生教育外部评估的权威性强、建设周期长,因此参评对象一般按照评估指标的内容来替代原先的发展规划,致使不同类型、不同层次的院校的建设方向呈现相似性。由此,评估主导部门通过评估制度影响乃至形塑了研究生教育的发展方向,利用标准化的指标体系实现对研究生教育质量的简化和统一,助推了参评对象同质性<sup>[9]</sup>。

外部评估不仅自身具有同质性生产的功能,而且它还通过评估指标的相似性来强化参评对象之间的同质化。评估指标的相似性首先体现为不同评估项目的指标体系在内容上高度重合,如学位授权审核、学位授权点合格评估、学科评估以及“双一流”建设成效评价的一级指标基本上都是围绕师资队伍、人才培养、教学条件、科学研究等维度,评估项目之间的指标差异性和针对性有待突出。其次,学科间的评估指标差异性不足。比较明显的表征是学位论文抽检中,某些省份的艺术学科不合格率(43.33%)远远超过平均不合格率(4.04%)<sup>[10]</sup>,而理工科等自然学科的合格率则长期处于前列<sup>[11]</sup>。自然科学和逻辑实证主义始终居于学科声誉的塔尖,其他学科的合法性只能通过模仿其方法和范式才能取得<sup>[12]</sup>。这意味着不仅相同类型的参评对象呈同质化的发展趋势,甚至不同性质的参评对象也朝着同质化方向发展。可见,评估文化及其附带的同质化基因在研究生教育中留下了深刻印记。

## (四) 评估的诊断改进功能较难发挥

美国学者斯塔弗尔比姆(D.L.Stufflebeam)就教育评价的目的有过经典论述。他强调:“评价最重要的意图不是为了证明,而是为了改进。”然而,我国研究生培养单位存在的缺陷难以在外部评估中得到体现和反馈。当前,研究生教育资源与评估之间建立了稳固的关系,如学科评估、“双一流”建设成效评价的结果成为我国政府财政拨款、学校发展规划的重要依据。这种“论功行赏”式的结果应用使外部评估成为一种利害性评估<sup>[13]</sup>。此外,外部评估需要向社会公开评估结果,即使是合格类的评估结果也会成为培养单位的质量符号和标签,从而影响院校的社会声誉和招生质量。例如,一些高校将《研究生教育发展质量年度报告》视为招生宣传。由此,外部评估在联系货币资本之外还具有了

象征性资本，无形中给参评对象造成压力。

为了在竞争中取得优势，高校倾向于凸显优势而非探究问题，这也导致外部评估难以发挥诊断与改进功能。在部分高校评估行动者的理解中，外部评估遵循竞赛思维，评估工作等同于“写本子”，需要通过文字功夫、数据拆分重组等“策略”来包装自己，以此获取与评估结果相对应的教育资源。例如，合格评估存在学位授权单位的自我评估深度不足、同行评审专家大多给予正面积积极的评价等问题<sup>[14]</sup>；学科评估中存在资源整合乱象，甚至是数据造假、学术不端等行为<sup>[15]</sup>。从客观上来说，适当参考评估结果能够为教育管理部门降低决策失误的风险，但当评估结果成为教育决策的主要工具时，评估则容易催生大学急功近利的不良风气。此外，评估组织者未能形成健全的反馈机制，在结果反馈上存在重告知轻改进的倾向<sup>[15](P111)</sup>。评估组织部门一般仅通过公告、通知等形式将结果反馈给参评对象，同时学科评估与专业学位水平评估虽然会给参评对象提供详细的反馈报告，但也只是告知其在各项指标上的排名情况，因此专家咨询的功能未能充分体现。

### 三、中国研究生教育外部评估体系问题的成因探析

由上可见，研究生教育外部评估中还存在一些问题，然而这些问题更多的是研究生教育外部评估体系中存在的表象问题，只有对上述问题表征的深层原因进行充分挖掘，才能更有针对性地构建外部评估体系的改革路径和调适措施。研究发现，以下几方面认识误区和实践偏差是影响评估有效性的重要原因。

#### （一）资源质量观主导评估标准

研究生教育外部评估的最终指向不是评估本身，而是研究生教育质量，因此评估活动的评价标准取决于对教育质量的界定和理解。阿斯汀（Astin）将质量观分为神秘的（mystical）、声誉的（reputational）、资源的（resources）、结果的（outcomes）和增值的（value-added）五种类型，其中后四种质量观已经有了相应的测量方法<sup>[16]</sup>。从质量观的多样化可以发现，教育质量不仅在定义上缺乏共识，而且在构成要素上也无法统一，要制定获得各方

认可的评价标准存在非常大的难度。现实情况是，不同群体的利益相关者选择有利于自身利益合法化的概念来评定研究生教育质量<sup>[17]</sup>。虽然政府评估行动者在我国评估活动中占据主导地位，但在评估标准的制定方面，他们通常听从学术带头人的意见，可以说高校人员把控着研究生教育评估的标准。在阿斯汀的五种质量观中，教育者倾向于采用资源质量观来定义质量，将师生比、生均经费、实验室规模等教育资源等同于教育质量，因为教育质量与资源建立起联系后，大学可以借助外部评估获取政府的财政支持，同时可以减轻自身在保障教育质量中的责任<sup>[18]</sup>。

随着质量观不断发展和演变，公众和政府对于结果质量观的呼声日益增长。由于教育产出具有不可视化、滞后性等特点，评估标准制定者开始使用科研绩效来代表教育成果，将教育产出转化为论文、项目、奖项等显性指标，然而国内外实证研究均显示科研产出与教学效果之间不具有正相关关系<sup>[19][20]</sup>。此外，声誉质量观在评估体系中也有所体现，如学科评估和专业学位水平评估增加了声誉评价。从现实情况来看，我国研究生培养单位的科研产出、社会声誉与教育投入之间息息相关，由此看来评价标准依然囿于资源质量观的框架中。资源质量观虽然极大地调动了高校领导者参与评估的积极性，但实际上难以真实反映研究生教育的成效，而且容易造成院校和学科之间的“马太效应”。更严重的是，评价体系过度依赖这类规模性指标后，教育评估开始迈入“五唯”陷阱，整个研究生教育系统也处于盲目追求“投入最大化”的境地。

#### （二）缺乏统筹规划和归口管理

研究生教育外部评估体系是由多项评估项目共同组成的整体，需要处理好系统整体和项目元素之间的关系。目前我国研究生教育外部评估项目类型众多，既有来自同一层级不同部门牵头组织的评估项目（如学科评估与专业学位水平评估由学位中心下的不同部门实施），也有来自国家、省（自治区、直辖市）等不同层级的项目（如博士、硕士学位点授权由国务院学位委员会和省级学位委员会分别负责）。从评估项目的组织单位来看，领导研究生教育评估的国家部门有国务院学位委员会及其办公室、国务院教育督导委员会及其办公室等，负责组织与落实的行政部门和专业组织包括教育部学位中心、教育评估院、国

务院学位委员会学科评议组和全国专业学位研究生教育指导委员会等。然而,由于统筹规划不足,从事研究生教育外部评估工作的部门缺少归口管理,评估主体各自为政,评估项目之间的数据不能共享,造成了研究生教育评估多头管理、评估内容重复交叉、评估数据多次提交的现象,评估任务往往呈现多、散、急的特点。这不仅人为增大了各方评估行动者的工作量,而且不同评估部门在数据采集时存在口径不一致的问题,相似的评估内容还可能出现结果相差较大的情况,由此导致部分参评对象质疑评估结果的真实性,并引发其对外部评估的反感情绪。

### (三) 技术理性主导评估设计

我国多数研究生教育外部评估部门代表着政府的意志,评估行动者将外部评估视为监测研究生教育质量的技术手段。在技术理性的视角下,评估人员必须将质量这一虚幻概念转化为可审计的标准化概念,其方法是通过规定清晰可识别的参与者、设计可预测结果的步骤、采用显性可量化的指标来简化教育质量的内涵<sup>[21]</sup>。利用技术视角来指导评估实践是有益且必要的,但倘若将评估仅仅视为一种技术过程,则具有明显的局限性。

第一,受技术理性指导的评估活动为了实现质量概念的标准化,倾向采用整齐划一的评价标准,强调参评对象的共同点、排斥其特异性<sup>[22]</sup>。这反映在外部评估中的表征,就是采用高度相似的指标体系评价各具特点的参评对象,导致质量只是等同于符合常规的标准,无法识别“出类”和“拔萃”的培养单位,最终抑制评价对象的多样性和特色化,造成“千校一面”的趋同现象。第二,以技术理性为设计理念的评估预设了不同利益相关者拥有一致的行动目标,并认为技术改进能够克服评估中面临的问题<sup>[23]</sup>。然而,评估系统和高校组织的复杂性意味着只有少数人具有共同目标,评估的技术理性观点往往掩盖了评估活动的行政属性。多数外部评估当中隐含着管理心态和控制形式,但评估的技术流程将其包装成平等的治理过程,这就忽视了参评对象的意愿,也容易掩盖高校内部的权力分配和资源竞争<sup>[24]</sup>。例如,学科评估中“不敢不参评”“资源整合”等现象便是行政问题的集中体现。近年来,我国研究生教育评估虽然在指标体系、实施过程等方面都较以往有了许多改进,并试图从方法和技术层面

处理好科学性、精准性、操作性、国际化、本土化这些关系<sup>[7](P70)</sup>,但评估体系中的深层矛盾仍难以破除。可见,技术理性也在一定程度上限制了评估行动者解决问题的思路。

### (四) 忽视高校教师的主体性

教师参与评估是保证评估效果和教学质量“黄金准则”<sup>[25]</sup>。教师在评估制度下的工作环境及其对制度的回应,与评估成效休戚相关。尽管我国高校教师参与了研究生教育评估活动,但外部评估的设计和執行往往是直接施加于高校教师的,普通教师几乎没有机会对评估目标、评价标准和评价流程提出意见,他们的作用仅限于向评估人员提供繁复数据,而评估的要求却最终传导至教师身上,并显著增加其工作压力<sup>[26]</sup>。这表明高校教师不仅在研究生教育评估的制度设计上没有话语权,而且在评估实施过程中也没有适应的余地,相反却需要接受评估对自身工作和职业生涯产生的重大影响。

有学者指出,“教师在评估中处于边缘化的位置”,当大学教师认为自己在评估体系中是“局外人”时,他们会产生被管理者 and 外部评估机构不重视或不信任的感觉<sup>[27]</sup>。已有研究发现,多数高校教师对评估活动持负面看法,教师们认为评估侵犯了自己的专业自主权,并且多数时候是一种管理活动而非质量提升活动<sup>[15](P109)[28]</sup>。外部评估一方面要求教师准备大量文件资料和填写表格,牺牲教师从事直接提高质量的活动时间;另一方面,区别对待教师的科研工作和教学工作,将教师的工作重心引向学术研究、偏离教书育人。这些做法会导致教师反感、批评甚至抵制外部评估。需要指明的是,评估本身并不产生质量,质量提升以一线教师主动投注精力为前提<sup>[29]</sup>。倘若外部评估活动无法与基层教师的质量改进工作建立联系,那么“以评促改”的效果不显著便成为必然结果。

## 四、中国研究生教育外部评估体系的完善路径

为了破除研究生教育外部评估体系中的问题,纠正其中的认识误区和实践偏差,未来可通过培育第三方评估机构、搭建研究生教育监测数据平台、强化研究生教育督导、开展研究生教育元评估来建成更加成熟有效的评估体系。

### （一）培育第三方评估机构，增强外部评估专业性

评估机构对研究生教育外部评估的设计与实施至关重要，其职责包括设计评估方案、遴选专家、构建评估指标体系、采集与整理数据、分析数据、协调与汇总专家意见、给出评估结果和撰写评估报告等。在“管办评分离”的高等教育管理体制下，独立的第三方评估机构理应成为研究生教育外部评估的主体。第三方评估机构一般围绕着特定的评估目标组建而成，具备专业的评估队伍和充足的评估时间，能够确保评估活动的专业性<sup>[30]</sup>。此外，独立的第三方评估机构具有高度自治性，与政府教育部门不存在隶属关系，能够避免出现政府部门兼任“运动员”与“裁判员”的情况，与高校也不存在利益相关关系，因而其评估结果的客观性和可信度更强。在研究生教育外部评估的改革实践中，我国亟需培育独立的第三方评估机构。一方面，政府作为研究生教育的宏观调控者，需有意识地培育多个第三方评估机构，可通过项目委托、责任授权、签订委托代理合约等形式，将外部评估项目让渡给社会上的第三方评估机构，使其获得代理人的独立身份，形成机构之间的多元特色和良性竞争，促使第三方评估机构不断完善服务功能。另一方面，第三方评估机构不能仅呈现最终的评估结果，也需要规范地说明评估标准和评估流程、加强信息公开，允许社会各界对其进行监督，从而提升自身的专业性与公信力。

### （二）推进研究生教育监测评估，减轻评估工作量

了解研究生培养单位的基本状态、收集数据信息是开展研究生教育评估的前提。研究生教育监测评估能够借助现代信息技术常态化地收集教育系统中关于研究生教育各要素的信息，包括学科建设、师资队伍、人才培养、科研情况等。评估部门可以根据需求进行数据挖掘与分析，从而判断教育系统中各单位研究生教育发展状况，方便指导其持续发展<sup>[31]</sup>。监测评估的核心在于加强教育评估的信息化，建立研究生教育评估的大数据平台。对于评估组织者来说，监测数据平台能够实现数据资源的开放共享和分级使用，在统一数据口径的情况下满足不同评估部门的诉求，提高不同评估项目的精确性，为统筹管理多个评估项目创造可能性。对于参评对象而言，监测评估有助于提高数据使用率，减

轻提交数据、重复收集的工作负担，同时避免评估“进校”影响教育教学日常工作。教育主管部门未来需依托高校、省域和国家三个层面，搭建研究生教育质量监测数据平台，利用区块链等信息技术优势，实现对研究生教育全过程的数据采集、存储和共享，推动研究生教育监测评估的深入开展。

### （三）强化研究生教育督导，凸显评估的诊断与改进功能

我国研究生教育外部评估存在诊断与改进功能难以发挥的问题，而研究生教育督导恰好能弥补这一缺点，既能够监督研究生教育活动的效果、反馈教育实际情况，又可以指导教育工作实践、调整教育生态体系<sup>[32]</sup>。具体而言，研究生教育督导可通过外部督导、内部督导两种形式来发挥诊断与指导功能。外部督导依托材料考察、实地调研、远程听课等方式，检查高校的研究生教育是否符合人才培养规律、是否执行相应的教育政策，指出培养单位存在何种问题，推动高校明确责任主体、加强原因分析和后续整改。与外部督导相比，内部督导机构能够洞察内部独特的教育生态，并进行深入的自我反省和自我修正。研究生培养单位内部的督导机构需要重视外部评估结果的分析与应用，对评估结果进行多类型、深层次的数据挖掘，履行内部自查、再评估、指导、咨询等职能，建立起内部质量保障和外部评估之间的联系，使研究生教育评估体系成为一个自律与他律并存的有机整体。

### （四）开展研究生教育元评估，提高评估品质

元评估是按照一定的标准或原则对评估工作本身进行评估的活动，其目的是判断评估工作的质量<sup>[33]</sup>。研究生教育元评估是加强研究生评估制度建设、完善评估工作的重要手段。我国研究生教育评估尚未进入成熟阶段，在评估项目的诸多要素上有待改进，积极开展元评估是适应新时代教育评价改革的现实要求。研究生教育元评估需描述教育评价的全过程，考察评估项目本身的科学性和有效性，对评估理念、评估技术和评估主客体的行为表现进行价值判断，帮助评估活动审视和反思自身，并为研究生教育评估体系提供激励、监督和指导。目前我国研究生教育元评估活动尚未发育成熟，未来首先需要建设符合中国研究生教育评估实际情况的元评估理论，明确元评估的目的、意义、作用和工作

重点;其次,要将元评估纳入研究生教育外部评估体系中,在研究生评估项目的设计方案中设置元评估程序,使其制度化;最后,应积极培育元评估组织,为研究生教育外部评估项目提供诊断报告,反馈教育评估活动中的真实问题及深层原因,从而形成外部评估体系的改进机制。

正如盖格指出:“评估不可能做到精确无误,评估结果也难以完全令人信服,要获得各方一致公认的评估结果则更难。”<sup>[34]</sup>由于教育系统和评估体系自身的复杂性,研究生教育外部评估不可避免地存在局限、面临争议。研究生教育外部评估是十分复杂的研究课题,如何发现评估问题、提升评估实效,仍然需要大量的思辨和实证研究。未来的研究生教育评估改革一方面可以依据已有研究成果,革除外部评估体系中的偏误,另一方面也应意识到外部评估仅是质量保障手段中的一种,不可过分放大和神化外部评估的作用,而应积极寻求外部评估与其他教育质量保障手段的协同和互补。

### 参考文献

- [1] 研究生教育评估制度研究及体系构建课题组.国外研究生教育评估制度研究[M].上海:华东师范大学出版社,2015:4.
- [2] 黄福涛.高等教育质量保障的国际趋势与中国的选择[J].北京大学教育评论,2010(01):114-124+191.
- [3] 包水梅.学科评估制度有效性:内涵特征与理论框架[J].教育科学,2021(06):62-67.
- [4] 么加利,罗琴.高等教育评价的数字依附及消解[J].高校教育管理,2022(01):26-37.
- [5] 俞蘂.大学评估何处去?国际评估在中国一流大学的兴起、扩散与制度化[J].华东师范大学学报(教育科学版),2022(01):30-42.
- [6] 靳玉乐,朴雪涛,赵婷婷,刘小强,司林波.笔谈:新时代教育评价改革与制度创新[J].大学教育科学,2021(01):13-25.
- [7] 陈学飞,叶祝弟,王英杰,熊庆年,龚放,胡建华,秦惠民,陈洪捷,施晓光,刘海峰,董云川,马陆亭,卢晓中,周川,樊秀娣,王鸿.中国式学科评估:问题与出路[J].探索与争鸣,2016(09).
- [8] 张德祥,朱艳.基于制度视角的大学组织发展同质化研究[J].教育科学,2011(06):61-65.
- [9] 李政.项目制下的高职院校同质化:作用机制、问题表征与改革路径[J].高校教育管理,2022(02):100-109.
- [10] 胡亮.艺术硕士学位论文质量监控及抽检评议机制研究[J].中国高等教育,2019(07):52-54.
- [11] 占茜,赵金敏.政策悖论框架下学位论文抽检政策的效能分析[J].江苏高教,2021(02):28-37.
- [12] Bloom A. The Closing of the American Mind[M]. New York: Simon and Schuster, 1987: 78.
- [13] 刘振天.中国高等教育评估体系及评估市场完善化[J].高教发展与评估,2014(04):19-28,117-118.
- [14] 贾金忠.学位授权点专项评估的特征、问题与对策研究[J].研究生教育研究,2016(03):72-75.
- [15] 张应强.“双一流”建设需要什么样的学科评估——基于学科评估元评估的思考[J].清华大学教育研究,2019(05):11-18.
- [16] Astin A. When Does A College Deserve to be Called High Quality[J]. Current Issues in Higher Education, 1980(01):1-9.
- [17] Martin M, Stella A. External Quality Assurance in Higher Education: Making Choices[M]. UNESCO, International Institute for Educational Planning, Paris, 2007: 30.
- [18] Skolnik M L. If the Cut Is So Deep, Where Is the Blood?: Problems in Research on the Effects of Financial Restraint[J]. The Review of Higher Education, 1986(04):435-455.
- [19] Hattie J, Marsh H W. The Relationship between Research and Teaching: A Meta-analysis[J]. Review of Educational Research, 1996(04):507-542.
- [20] 李宝斌,许晓东.高校教师评价中教学科研失衡的实证与反思[J].高等工程教育研究,2011(02):76-81+123.
- [21] Blanco Ram í rez, Gerardo. Studying Quality beyond Technical Rationality: Political and Symbolic Perspectives[J]. Quality in Higher Education, 2013(02):126-141.
- [22] Bensimon E M. Total Quality Management in the Academy: A Rebellious Reading[J]. Harvard Educational Review, 1995(04):593-612.
- [23] Lee Harvey, Jethro Newton. Transforming Quality Evaluation[J]. Quality in Higher Education, 2004(02):149-165.
- [24] 张应强,苏永建.高等教育质量保障:反思、批判与变革[J].教育研究,2014(05):19-27+49.
- [25] Pat Hutchings. Opening Doors to Faculty Involvement in Assessment.[EB/OL]. (2010-04)[2022-02-08]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.450.9679&rep=rep1&type=pdf>.
- [26] 蔡红红,王思遥.评估结果何以影响学科后续建设成效——工作资源与教师工作要求的中介作用[J].高等教育研究,2022(01):37-47.
- [27] Hoecht A. Quality Assurance in UK Higher Education: Issues of Trust, Control, Professional Autonomy and Accountability[J]. Higher Education, 2006(04):541-563.
- [28] Anderson G. Assuring Quality/Resisting Quality Assurance: Academics' Responses to 'Quality' in Some Australian Universities[J]. Quality in Higher Education, 2006(02):161-173.
- [29] Newton J. Feeding the Beast or Improving Quality?: Academics' Perceptions of Quality Assurance and Quality Monitoring[J]. Quality in Higher Education, 2000(02):153-163.
- [30] 耿宗程,王和平.比较视野下我国高等教育第三方评估的

- 公信力困境及其超越[J].高教探索,2022(01):22-28.
- [31] 唐丽.基于区块链的研究生教育监测评估模式[J].现代教育管理,2019(09):113-117.
- [32] 周文辉,曹镇奎.高等教育督导评估:内涵特征、内在冲突和路径选择[J].江苏高教,2019(05):28-34.
- [33] 张荣娟,徐魁鸿.美国高等教育元评估制度探析——以高等教育认证委员会为例[J].高教探索,2018(02):65-69.
- [34] [美]盖格.研究与相关知识——第二次世界大战以来的美国研究型大学[M].张斌贤,译.保定:河北大学出版社,2008:8.

## External Evaluation System of Postgraduate Education in China: Policy Review, Problem Analysis and Improvement Path

ZHOU Wen-hui ZHAO Jin-min

**Abstract:** The evaluation system of postgraduate education in China is characterized by the localization dominated by external evaluation, and a scientific external evaluation system is very important for the high-quality development of graduate education. After 40 years of exploration and reform, the external evaluation system of postgraduate education in China has achieved a certain effect in the system construction, but there are still practical problems that need to be improved. The specific manifestations include: the scientific nature of the evaluation index still needs to be improved, the collection of evaluation data increases the workload of colleges and universities, the evaluation leads to the homogeneity of the participants, and the evaluation's diagnostic improvement function is difficult to play. The main reasons for these problems are the resource quality view dominates the evaluation criteria, the lack of overall planning and centralized management, the technical rationality dominates the evaluation design, and the neglect of the subjectivity of college teachers. To meet the reform needs of educational evaluation in the new era, it is necessary to cultivate third-party evaluation institutions, build monitoring data platforms, strengthen the supervision of graduate education, and carry out meta-evaluation, so as to build a more mature external evaluation system of graduate education.

**Key words:** postgraduate education; external evaluation system; policy review; third party assessment; monitoring and evaluation; meta-evaluation

(责任编辑 李震声)