

省级政府高等教育评价改革政策内容 再生产及其影响机制研究^{*}

——基于机器学习与定性比较分析的实证研究

陆根书^{1,2} 董宇婧¹ 刘扬云³

(1. 西安交通大学人文学院, 西安 710049; 2. 西安交通大学中国西部高等教育评估中心, 西安 710049;
3. 中国教育发展战略学会教育评价专业委员会, 北京 100084)

摘 要: 高等教育评价改革是增强中国高等教育综合实力和建设高等教育强国的关键环节, 省级政府在制定高等教育评价改革政策的过程中对中央政策内容的再生产对当地高等教育评价改革实践将产生重要影响。通过机器学习构建了省级政府政策内容再生产系数, 对我国 31 个省级政府的高等教育评价改革政策内容再生产程度进行了测算, 并进一步运用模糊集定性比较分析方法 (fsQCA) 揭示了省级政府高等教育评价改革政策内容再生产的影响机制。研究发现: 省级政府高等教育评价改革政策内容再生产程度总体偏低。省级政府高等教育评价改革政策内容再生产程度高的组态有五种, 可进一步概括为三类路径组合, 即压力传导型、横向竞争型和自主优化型。研究也发现省级政府政策内容再生产程度不高的组态有三种, 可进一步概括为二类路径组合, 即资源缺乏型和压力异化型。通过对比政策内容再生产程度高与不高的路径发现, 省级政府教育财政投入是省级政府推进高等教育评价改革政策内容再生产的经济基础, 省高等教育发展综合实力是省级政府推进高等教育评价改革政策内容再生产的现实基础。省级政府教育财政投入与省级政府的注意力配置存在替代效应, 二者均可以对外部压力进行积极回应; 不存在外部压力时, 其与省高等教育发展综合实力存在协同作用, 二者的配合可以有效提升省级政府的政策内容再生产程度。因此, 省级政府应激发内生动力, 合理配置教育财政投入, 因地制宜地贯彻落实中央政策; 中央政府则应进一步加强督导检查及奖惩力度, 以更好地推动相关政策落实落地。

关键词: 高等教育评价改革政策; 政策内容再生产; 机器学习; fsQCA

一、引言

高等教育评价改革肩负着构建中国特色高等教育评价体系、建设高等教育强国、实现高等教育现代化的历史重任和时代使命(刘振天, 2024)。2020 年 10 月, 中共中央、国务院印发了《深化新时代教育评价改革总体方案》(以下简称“《总体方案》”), 对教育评价改革工作做出战略部署。此后, 中央政府层面相继出台《关于正确认识和规范使用高校人才称号的若干意见》《关于破除高校哲学社会科学评价中“唯论文”不良导向的若干意见》等政策文件, 形成了高等教育评价改革的顶层政策设计。在我国当前的治理体系下, 地方政府对高等教育评价改革的治理过程和政策实践依赖于中央的顶层设计, 但作为“文件传递链”起点的中央政策文件因需要统合不同地区的治理情境, 常侧重于方向性、原则性和指导性的宏观叙事, 模糊性成为中央政策“不可剥离的属性”(薛增鑫, 满小欧, 2024), 这就意味着地方政府需要对中央文件进行政策内容再生产, 即基于中央发布的政策文件精神对本地区相应的中

^{*} 基金项目: 国家社会科学基金教育学一般项目: “大学毕业生教育-职业错配及其收入效应: 多维视角的研究”(BIA240161)。

央政策进行细化和更新(刘河庆,梁玉成,2021)。这一过程也是将中央政策意图转化为地方具体治理效能的关键环节。可见,地方政府的高等教育评价改革政策内容再生产行为隐含了其对中央政策的精准回应,不仅是当前我国治理体系下政策落地的必要手段,也是避免中央高等教育评价改革政策目标在逐级传递中发生“偏移性风险”的重要基础。但是,在高等教育评价改革文件传递过程中,部分地方政府出现了只是充当文件“二传手”的现象,发布的高等教育评价改革政策与中央政策存在内容高度同质化问题,能动性与创新性不足(秦一帆,2024)。这种“以文件落实文件”将在很大程度上影响地方政府政策执行的效率(李瑞昌,2012),从而影响国家高等教育评价改革政策的治理效能。

为了解省级政府在贯彻落实中央高等教育评价改革政策的状况及机制,本文提出了如下研究问题:省级政府在对中央政府发布的高等教育评价改革政策的细化和更新程度上有何差异?这种差异会影响政策实践效果吗?省级政府充当中央文件“二传手”或创新先行者的机制是什么?有哪些因素发挥了关键作用?

目前学界对于高等教育评价改革政策的研究主要集中于价值探寻与实践反思两方面,虽然对省级政府的高等教育评价改革政策实践有所关注,但未考虑到省级政府的异质性,且以往采用的一些定性阅读和文本计量分析方法很难对政策内容细化与更新程度进行准确测量。有鉴于此,本研究聚焦省级政府对高等教育评价改革政策的内容再生产,通过机器学习测算政策内容再生产程度,并进一步运用模糊集定性比较分析方法考察政策内容再生产的影响机制,以期为贯彻落实中央高等教育评价改革政策并提高地方政策的适切性提供参考。

二、高等教育评价改革政策研究概述

高等教育评价改革是现代高等教育治理的重要环节(周洪宇,2020),高等教育评价改革的政策文本则是政府高等教育评价改革理念的具象载体,也是政府推动高等教育评价改革行动的重要起点。目前学界对高等教育评价改革政策的研究主要遵循两条路径:

一是价值探寻。主要从宏观层面探析我国高等教育评价改革政策的价值导向,侧重于解读高等教育评价改革政策的理论内涵。过去根植于高等教育评价中的“五唯”困境是遵循效率化、标准化、方法化、规模化和脱主体化的价值导向(易凌云,2021),虽然“五唯”评价最大限度回应了在高等教育大众化时代国家和社会对教育效益和标准化人才供给的诉求,但内嵌的功利主义思想忽视了人与高等教育的长远发展。近年来,高等教育向高质量发展跃迁,以人为本和教育公平理论重回高等教育场域,立德树人和政学结合成为建构新时代中国特色高等教育评价体系的价值逻辑(周光礼,2023)。具体而言,一是要引导高校聚焦立德树人,将立德树人的成效作为检验学校一切工作的根本标准,要立以人才培养指标为主的评价、立以过程评价和增值评价为主的方式,并将师德师风作为评价教师队伍素质的第一标准;二是要引导高校服务国家战略需求,坚持分类指导、分类评价原则,通过评价指挥棒引导高校牢牢把握重要使命,全力服务新时代多样化需求。可见,关于高等教育评价改革政策的价值研究对“为什么要进行高等教育评价改革?”以及“需要改什么”等问题进行了讨论,但研究还停留在对《总体方案》等政策的内容解读与价值内涵分析上,对配套政策以及政策落地过程关注不足,未能回应政策价值内涵如何转化为实际行动这一现实问题。

二是实践反思。聚焦于政策过程,主要探讨我国高等教育评价改革政策的发展过程及中央和地方政府在政策制订与实施过程中的配合。对政策发展过程的分析,主要通过政策文本内容回溯高等教育评价改革政策的演进轨迹,审视政策变迁规律。对于我国高等教育评价改革政策变迁的研究表明,政府在高等教育评价改革中始终发挥着主导作用,但过去的研究多关注中央政府在政策顶层设计中的历时性行为逻辑,并没有将地方政府的政策纳入分析范畴。有鉴于此,近年来有关高等教育评价改革政策的研究逐渐拓展至省级政府层面的政策实践。研究者聚焦于政策采纳过程,剖析政策执行的阻滞因素(阳荣威,刘伟豪,2023)。中央政府制定的高等教育评价改革政策不可能全面考虑不同地区的差异

性,地方政府在执行相关政策时需要从地方实际出发,因地制宜地对中央指导性政策进行政策内容再生产,以实现地方政策与中央顶层设计的统合性,以及和地方实际的适恰性(林雪霏,2015)。但在高等教育评价改革政策的实际执行过程中,一些地方政府虽然响应快速,却仅发挥了政策“上传下达”的功能,并没有根据地方实际进行政策内容的再生产,政策执行的地方针对性不足,效率低下(秦一帆,2024)。

由上可见,学界对高等教育评价改革政策虽然进行了一些研究,对推进政策完善与落实做出了积极贡献,但仍有不足之处,主要表现在两个方面:一是现有研究主要是对《总体方案》的解读,对地方政策实践关注不足,对地方政府政策制定行为还没有形成系统认识,对“中央文件-地方行动”的逻辑链条研究还没有形成闭环。二是虽然有少数学者对地方政府的高等教育评价改革政策有所关注,但对地方政府的特殊性与异质性未充分考虑。此外,囿于方法的局限性,以往采用的一些定性阅读和文本计量分析方法难以对政策内容再生产程度进行准确测量,使得已有研究仅关注了地方政府的政策采纳速度、政策名称以及政策表述形式等行为,还没有对地方政府高等教育评价改革政策内容进行分析。因此,从政策内容再生产视角分析我国地方政府对中央高等教育评价改革政策的响应程度及影响机制,有助于深入理解我国地方政府对中央高等教育评价改革政策采纳与执行的内在逻辑。

对政策内容再生产的研究始于学界对政策采纳的关注。政策采纳是指一项公共政策被不同政府主体采纳和应用的动态演进过程(Shipan & Volden, 2006)。在中国以科层治理为特征的体系下,中央政府发布政策文件后,作为政策重要传递者的省级政府必然会做出响应,而省级政府的政策采纳行为是一种直观的政策响应(易兰丽,范梓腾,2022)。一些研究者出于易于量化的考虑,常把省级政府是否采纳和采纳速度作为其政策采纳行为的测量指标,但这种“表层”的测量不够深入、准确,而且也暗含着“采纳速度越快越好”的价值取向,对省级政府政策采纳的内容和地方治理情境“适切性”的关注不足。基于此,一些对政策采纳的研究开始转向“深层”测量,关注省级政府对政策的借用、再造、扩张和内容再生产等。其中,政策借用、政策再造和政策扩张主要涉及同级政府间的政策传递与政策学习(Jansa et al., 2019; Chen & Huang, 2021),政策内容再生产则关注政策文件传递链中下级政府对上级政府政策文件的细化和更新程度(刘河庆,梁玉成,2021),强调对政策内容和要素的操作化设计和地方性知识的嵌入(仇叶,2021)。近年来,随着计算社会科学的兴起,文本挖掘技术为政策内容再生产程度的测量提供了方法支持,“文本相似度”成为测量政策内容再生产程度的一个有效指标(Hinkle, 2015)。

政策内容再生产本质上是政策创新扩散过程中的政策采纳行为(张晓君,王伟桥,2024),因此,关于政策内容再生产影响因素的研究也多以政策创新扩散理论为基础。政策创新扩散理论对政策采纳的影响因素分析最初存在两种模型:内部决定因素模型和区域扩散模型。内部决定因素模型认为,政府内部及其管辖范围内的经济、社会和政治情境是影响政策采纳的决定性因素;区域扩散模型则认为政策扩散过程会受到地理位置等因素的影响,强调外部因素的重要性(郭伟,2021)。Berry等人认为两种模型可以在理论上实现统一,其将两种模型结合,提出了政策创新扩散总模型,认为地方政府的政策采纳行为会受到内外部因素的共同影响(Berry & Berry, 1990)。其中,内部影响因素包括政治、经济、文化以及领导者特性等,外部影响因素包括水平因素和垂直因素,水平因素主要涉及邻近政府间的竞争压力和学习动力,垂直因素是指上级政府的强制或激励(Volden et al., 2008; 王洪涛,陈洪侠,2017)。一些学者研究省级政府政策内容再生产时,借鉴了政策创新扩散总模型,重点关注影响政策内容再生产的内外部因素,并结合具体情境对内外部因素进行了操作化处理,其中内部因素主要涉及地方政府政策偏好、财政能力、人口结构、经济发展水平、市场化程度等,外部因素主要涉及同级政府的学习、竞争以及中央政府的行政压力等(刘河庆,梁玉成,2021; 薛增鑫,满小欧,2024; 张晓君,王伟桥,2024)。此外,政策创新扩散理论在高等教育领域也得到了应用,一些研究者从内外部视角探讨了高等教育领域相关政策创新扩散的影响因素,将内部因素操作化为地区经济发展水平、教育财政支出、地方财政压力、在校生规模、高职院校规模、大学规模、院校类型、第三产业规模等,将外部因素操作化

为同级横向政府间采纳比例、邻近地区政府间采纳比例、中央政策信号、邻里效应和社会公众关注度等(吴柯豫, 魏署光, 2025; 凌磊, 2018)。

由上可见, 已有一些研究探讨了地方政府政策内容再生产的内涵与影响机制, 为理解地方政府对中央政策的差异化响应提供了有益借鉴。然而, 由于已有研究选取的政策主题比较宽泛, 尚未形成共识性的分析框架和相对稳定的研究结论, 且对高等教育评价改革政策这一政策议题的关注较少。此外, 已有研究对隐含的“政策内容再生产程度越高越好”的假设也没有进行验证, 且关于政策内容再生产影响机制的研究主要依托传统的量化分析范式, 鲜少考察政策采纳的内外影响因素与政策内容再生产的非对称性因果关系^①以及各因素的相对重要性。因此, 本研究以中央政府和省级政府发布的高等教育评价改革政策文本为研究对象, 应用机器学习与定性比较分析方法, 探究省级政府高等教育评价改革政策内容再生产的逻辑, 以及受内外部因素的影响, 以期对我国高等教育评价改革政策的落实与优化有所裨益。

三、研究设计

(一) 分析框架设计

本研究借鉴政策创新扩散总模型的分析视角, 从内部与外部两个层面对省级政府高等教育评价改革政策内容再生产的影响因素进行分析。研究的具体分析框架如图 1 所示。

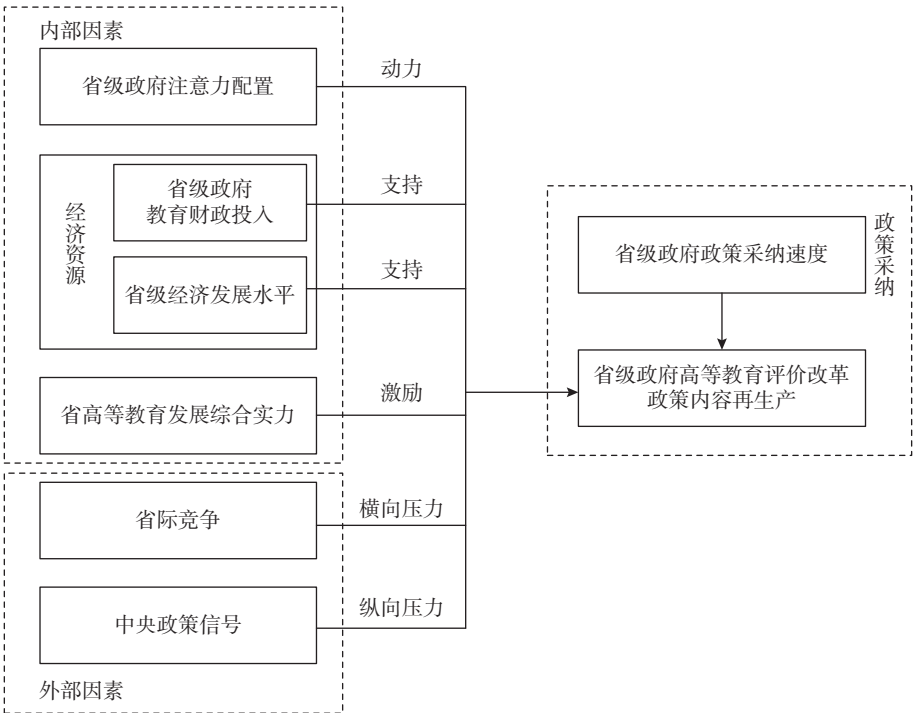


图 1 高等教育评价改革政策内容再生产的影响因素模型

就内部影响因素而言, 主要关注省级行政区域内的政治、经济和社会等方面因素对省级政府政策内容再生产行为的影响(Berry & Berry, 1990)。已有研究常使用市场化指数、人口规模、政策类型等指标进行分析, 但高等教育评价改革政策具有特殊性, 仅仅使用这类指标难以反映与该政策相关的地区异质性。本研究在已有研究的基础上, 根据高等教育评价改革政策特性, 从省级政府注意力配置、经济资源以及省高等教育发展综合实力等方面对内部影响因素进行量化分析。省级政府注意力配置的实质是选择性(Yantis, 2000), 特别是在“有限理性”约束和“多任务-多委托”的制度环境下, 省级政府

需要采取“串行处理”策略(Augier, 2002),其注意力在不同政策上的分配将会影响其资源分配(易兰丽,范梓腾,2022)。由于高等教育评价改革政策内容再生产需要省级政府在人力、财力和物力资源上的投入,其经济资源可以为政策内容再生产提供经济支持(李欢欢,顾丽梅,2020),具体包括教育财政投入能力和经济发展水平(张晓君,王伟桥,2024)。此外,省级政府在采纳中央政策并对其进行内容再生产前需要结合地区发展现状,对政策实施效果进行预估,能否在有限资源约束下较快取得明显成效是影响省级政府采纳政策并进行内容再生产的关键因素(薛阳等,2023)。因此,《总体方案》发布前各省的高等教育发展综合实力会影响省级政府的高等教育评价改革政策内容再生产程度。

就外部影响因素而言,影响省级政府政策内容再生产的外部压力涉及中央政府的要求、省际竞争与公众参与等。由于高等教育评价改革政策涉及公共意识形态塑造,具有省级政府对其关注度高,而相关压力群体与政府联系松散的特征(任锋,朱旭峰,2010),公众参与有待进一步提升(张辉蓉,盛雅琦,2022)。由于这个原因,本研究暂时不考虑公众参与,仅从中央政府的要求和省际竞争两个方面对外部因素的影响进行剖析。已有的一些研究将中央政府的要求操作化处理为中央行政指令和政策信号等指标(张海柱,林华旌,2022),由于中央政府的高等教育评价改革政策面向所有省级政府,就政策本身而言,省级政府接收的行政指令不存在异质性,因此本研究仅以中央政策信号来测量中央政府的要求。省级政府在贯彻落实中央政策时,也存在一定程度的竞争,本研究用省际竞争指标来描述各省之间的政策竞争。

除内外部因素外,本研究将政策采纳速度也纳入了理论模型,主要基于以下考量:首先,政策创新扩散总模型最初探讨的是内外部因素如何影响地方政府政策采纳行为(Berry & Berry, 2018),加上现有研究对于政策采纳的测量多聚焦于采纳速度等浅层特征(Mallinson, 2016),而政策内容再生产属于更深层次的采纳行为。这使得内外部因素对政策内容再生产的影响相较于浅层采纳行为可能存在着明显的差异(刘河庆,梁玉成,2021)。因此,在构建政策内容再生产的影响因素模型时,应该同时考虑采纳速度这一变量。这不仅有助于更全面的探究政策内容再生产的影响机制,也有助于重新审视以往研究所暗含的“政策采纳速度越快越好”的线性假设。其次,一些研究也发现,表征政策表层采纳行为的“采纳速度”会对深层采纳行为的“政策内容再生产”产生影响(陈刚等,2023),但对二者关系的研究尚未取得一致意见,仍有进一步探究的必要。

(二) 研究方法

本研究要解决的核心问题是省级政府高等教育评价改革政策内容再生产存在差异的程度及其影响机制。因此,首先要对省级政府政策内容再生产程度进行测量。为此,本研究选择 Doc2vec 模型,根据省级政府政策文本与中央政府政策文本的相似度对省级政府政策内容再生产程度进行测量(王法硕,张嘉玲,2023)。其次,运用定性比较分析(QCA)方法对省级政府政策内容再生产的影响因素进行组态分析。

1.Doc2vec 模型

本研究根据省级政府与中央政府的政策文本的相似度来测量省级政府政策内容再生产程度(Hinkle, 2015)。当特定省级政府发布的政策文件与中央政策文件的文本相似度越高,则表明省级政府对中央政策文本的细化和更新程度越小,即政策内容再生产程度越低。计算文本相似度需要将自然语言转换为计算机可识别的向量形式,再通过计算向量之间的距离得出最终结果,常用模型包括词袋模型、Word2vec 模型以及 Doc2vec 模型等。由于本研究探讨的高等教育评价改革政策均为内容详尽的长文档,因此选择擅长处理长文档和段落的 Doc2vec 模型进行分析(陈洁,2023)。该模型不仅能够保留语义之间的关系,而且其无监督算法能够最大程度地减少人工干预,从而规避关键词共现方法存在的规范性不足等问题(阮光册,夏磊,2019)。Doc2vec 模型又包括 PV-DM 和 PV-DBOW 两个子模型(杨宇婷等,2016),其中利用段落向量和词向量相结合方法来预测下一个词向量的 PV-DM 模型被证实

多数自然语言处理任务中的表现更好(Le & Mikolov, 2014, pp.1188-1196)。因此,本研究选择 Doc2vec 模型中的 PV-DM 模型进行文本量化分析,以向量之间的余弦相似度来计算文本相似度。

2.定性比较分析(Qualitative Comparative Analysis, QCA)

本研究选择定性比较分析方法,从组态视角分析省级政府在高等教育评价改革政策内容再生产方面存在的差异及其原因。定性比较分析方法是一种整合定性与定量研究取向的方法,通过分析多因素组合路径来解释复杂因果关系,广泛应用于社会科学及跨学科领域。其核心是基于集合论和组态思维,识别前因条件的不同组合如何导致特定结果。选择这一方法主要出于以下几点考虑:

第一,以 31 个省级政府的政策文本作为分析对象,样本量较小,使用传统回归分析方法可能导致模型结果不稳定,从而降低研究结论的可靠性,而定性比较分析方法更适合分析中小样本,并能对定性资料进行定量分析。

第二,虽然可以像已往一些研究那样将有关变量进行量化操作处理,但在我国当前的治理情景中,各变量之间难以实现完全相互独立,而定性比较分析方法关注变量之间的相互依赖以及多案例的“并发因果关系”,因此更适合从整体视角来解析省级政府高等教育评价改革政策内容再生产的影响机制,并对有关模式进行概括总结。

第三,在我国当前的治理情境中,省级政府的高等教育评价改革政策实践高度依赖于中央的顶层设计。各省在对高等教育评价改革政策进行内容再生产时可能存在结果相似性,这就意味着可能存在多条路径导向同一结果的“等效”因果链,使用传统分析方法难以刻画这种等效关系的异质性,而定性比较分析方法可以识别不同条件组态与被解释结果之间的等效性,并通过横向对比呈现因素的相对重要性(Fiss, 2011)。

定性比较分析方法有三种具体应用模式,即清晰集(csQCA)、模糊集(fsQCA)与多值集(mvQCA)。在本研究的样本数据中,结果变量和条件变量均为连续变量,难以明确地按照“是否属于”特定集合的标准进行划分,故采用能够处理部分隶属关系的模糊集定性比较分析(fsQCA)方法进行实证分析。

(三)数据来源

《总体方案》将教育评价改革提升至国家顶层设计的高度,教育部、国务院教育督导委员会、人力资源社会保障部等相关部门围绕高等教育评价改革政策内容发布一系列相关文件,并持续推进政策落实。本研究以中国人民大学评价研究中心发布的教育评价改革文件汇编为参考,遵循省级政府响应程度高、涉及面广的原则,在中央和各省级政府网站及相关职能部门官方网站中检索高等教育评价改革、高等教育等关键词,共搜集 2020-2024 年中央政府及有关部委发布的高等教育评价改革政策文件 12 份^②(表 1),31 个省级政府(不包括港澳台地区)对应出台的地方性政策文件 236 份,共计 248 份政策文件,本研究以此作为省级政府高等教育评价改革政策内容再生产程度测算的原始数据。此外,省级政府注意力配置数据及中央政策信号数据来源于各省级政府门户网站,其余变量来自 2020—2024 年的《中国统计年鉴》和《中国教育经费统计年鉴》。

(四)变量测量

1.结果变量

本研究关注的结果变量是省级政府高等教育评价改革政策内容再生产程度。本研究通过计算各省级政府政策文本与中央政策文本的相似度(刘河庆,梁玉成,2021),构建省级政府政策内容再生产系数(即 1-各省级政府政策文本与中央政策文本的相似度)来测量政策内容再生产程度。计算的具体步骤包括文本预处理、数据训练以及相似度计算。首先,使用 Python 编程的 jieba 分词工具,根据自定义词典和哈尔滨工业大学停用词表对政策文本进行预处理。其次,在 Python3.8 和 Gensim4.3.3 的实验环境下,以清华大学公开中文文本数据集 THUCNews 中时政、社会和教育文本资料作为训练数据,将文

档向量维度设置为 300, 训练窗口大小设置为 8(王法硕, 张嘉玲, 2023), 生成 Doc2vec 模型, 以此模型对 248 份政策文件进行向量化处理。最后, 以向量之间的余弦相似度来表示省级政府政策文件与中央政策文件的相似程度。经过处理后, 将省级政府政策内容再生产系数作为结果变量。

表 1 2020—2024 年中央政府发布的涉及高等教育评价改革的政策文件

编号	发文主体	政策名称
1	中共中央 国务院	深化新时代教育评价改革总体方案
2	中共中央办公厅 国务院办公厅	关于深化新时代教育督导体制机制改革的意见
3	国务院教育督导委员会	教育督导问责办法
4	教育部	关于正确认识和规范使用高校人才称号的若干意见
5	教育部	关于破除高校哲学社会科学研究评价中“唯论文”不良导向的若干意见
6	人力资源社会保障部 教育部	关于深化高等学校教师职称制度改革的指导意见
7	教育部 中央组织部 中央宣传部 财政部 人力资源社会保障部 住房和城乡建设部	关于加强新时代高校教师队伍建设的指导意见
8	教育部	本科毕业论文(设计)抽检办法(试行)
9	教育部 财政部 国家发展改革委	关于深入推进世界一流大学和一流学科建设的若干意见
10	教育部	普通高等学校本科教育教学审核评估实施方案(2021—2025年)
11	教育部 财政部 国家发展改革委	“双一流”建设成效评价办法(试行)
12	国务院办公厅	关于进一步做好高校毕业生等青年就业创业工作的通知

2. 条件变量

根据前面构建的分析框架, 本研究的条件变量分为三类: 第一类为省级政府注意力配置、经济资源和省高等教育发展综合实力等反映省级政府所在区域特征的因素; 第二类为以省际关系为核心的外部因素; 第三类为省级政府政策采纳速度。这三类条件变量的具体测量方式如下:

(1)省级政府注意力配置。省级政府工作报告可以直观反映省级政府一年的工作计划, 在一定程度上是其注意力配置的映射(邓雪琳, 2015)。参考已有研究(刘河庆, 2020), 本研究对各省级政府 2020—2024 年政府工作报告^③进行 LDA 主题模型分析^④。具体操作步骤为: 首先, 使用 Python 编程的 jieba 分词工具, 根据自定义词典和哈尔滨工业大学停用词表对省级政府工作报告进行预处理; 其次, 结合困惑度指标确定对于省级政府工作报告这一文本集合而言的最优主题数量(黄荣贵, 2017; Blei et al., 2003); 再次, 在此基础上使用 LDA 主题模型对省级政府工作报告进行分析, 识别涉及高等教育的主题, 以该主题出现的概率来测量省级政府对高等教育主题的注意力配置强度。

(2)省级政府教育财政投入。省级政府投入高等教育的财政经费会在一定程度上影响高等教育评价改革可动用的资源。本研究参考已有研究(吴柯豫, 魏道光, 2025; 吴峰, 石益静, 2024), 将省级政府教育财政投入操作化处理为 2019—2023 年一般公共预算内高等教育事业费和基本公共建设支出的年均值。

(3)省级经济发展水平。经济发展水平反映了各省的经济能力, 经济发展水平越高, 政府采纳新政策并创新的行为越频繁(Walker, 1969)。本研究将经济发展水平操作化处理为 2019—2023 年各省份人均 GDP 的年均数值(陈刚等, 2023)。

(4)省高等教育发展综合实力。参考已有研究(邵剑耀, 2024), 本研究应用各省份 2019 年高等教育地区竞争力指标来反映其进行高等教育评价改革之前的高等教育发展综合实力。当一个省 2019 年高等教育发展综合实力越强时, 一方面表明省级政府推进高等教育评价改革政策内容再生产的基础越好, 另一方面也表明该省更可能为了维持发展优势而积极推进高等教育评价改革政策内容再生产。

(5)省际竞争。本研究将省际竞争操作化处理为某个省级政府制定某项高等教育评价改革政策

时,已经发布了该项政策的省级政府数量。由于每个省份涉及到多项政策文件,因此选取均值作为该省份省际竞争的测量结果(陈刚等, 2023)。

(6)中央政策信号。政策信号表现为中央政府向下传达的一种引导或激励性信息,通常出现在中央政府召开的工作会议或者领导人的公开讲话中(张海柱, 林华旌, 2022)。对中央政策信号变量的操作化处理过程为:对习近平总书记 2020—2024 年间在各个省份考察时发表的重要讲话进行 LDA 主题分析,以涉及高等教育的主题出现概率作为识别中央政策信号强度的指标。与已有研究的测量方法相比,该指标能更好地体现中央政策信号的区域差异性。

(7)省级政府政策采纳速度。本研究采用如下方法对省级政府政策采纳速度进行测量:将中央发布特定高等教育评价改革政策的时间与各省级政府发布对应政策的时间进行比较,计算二者间隔的天数,再将其换算成以年为单位的政策采纳速度(刘河庆, 梁玉成, 2021)。由于同一省份存在多份政策文本的情况,本研究取均值作为该省份的政策采纳速度。该指标数值越大,表明省级政府发布对应政策的速度越慢。

3.变量校准

在确定结果变量与条件变量后,需要对变量进行校准。校准是给案例赋予集合隶属度的过程,根据变量类型,本研究采用直接校准法,分别选取变量的百分位值 95%(完全隶属)、50%(交叉点)和 5%(完全不隶属)作为定性锚点(Ragin, 2009)。利用 fsQCA 软件进行校准,将原始数据转换为介于 0~1 之间的模糊隶属分数,以此作为后续进行 fsQCA 分析的数据基础。为了避免条件变量隶属度恰好为 0.500 的案例在分析过程中被排除,本研究在数据处理时将 0.500 的隶属度调整为 0.499(Crilly et al., 2012; 杜运周等, 2022)。变量测量及校准如表 2 所示。

表 2 变量测量与校准

变量名称	测度标准	变量校准			数据来源
		完全隶属	交叉点	完全不隶属	
省级政府注意力配置	省政府工作报告中高等教育主题的出现概率	0.734	0.146	0.001	省级政府官网
省级政府教育财政投入	一般公共预算内教育事业费和基本建设支出(年均)	32 774 127.300	14 216 174.800	3 013 165.100	《中国教育经费统计年鉴》
省级经济发展水平	省人均GDP(年均)	153 365.900	63 567.600	47 139.100	《中国统计年鉴》
省高等教育发展综合实力	高等教育地区竞争力(2019年)	81.115	55.000	19.890	科教评价网
省际竞争	某省级政府制定特定政策时,已经制定该政策的省级政府数量	16.667	11.143	7.311	省级政府官网
中央政策信号	习近平总书记考察各省份时发表讲话中涉及高等教育主题的出现概率	0.999	0.535	0.001	省级政府官网
省级政府政策采纳速度	省政府发布政策与中央发布政策相差的具体时间(以年为单位)	1.483	1.079	0.402	省级政府官网
省级政府政策内容再生产系数	1-省级政府政策文本与中央政策文本的相似度	0.280	0.181	0.039	Doc2vec模型测量结果

四、实证分析结果

(一)省级政府高等教育评价改革政策内容再生产程度分析

根据前述方法,本研究对 31 个省份的高等教育评价改革政策内容再生产系数进行了计算,结果如图 2 所示。整体而言,省级政府高等教育评价改革政策内容再生产系数均值为 0.159,江西、浙江、黑龙江、陕西、江苏和山东等地的政策内容再生产程度较高,但约 70% 的省级政府发布的高等教育评价改革政策文件与中央政府出台的政策文件内容相似度超过 80%,省级政府政策与中央政策存在较高一

致性,多数省级政府发布的高等教育评价改革政策文件具有内容同质化的特征。

图 3 比较了不同地区省级政府对高等教育评价改革政策的内容再生产情况。从均值来看,省级政府高等教育评价改革政策内容再生产程度存在地域差异,中部地区的省级政府高等教育评价改革政策内容再生产系数均值最高,东部地区的均值次之,西部地区的均值最低。从箱线图的上下边界差距来看,各地区的内部差异性较为明显,政策内容再生产系数的跨度较大,表明即使处于相似地理条件下,各省份的高等教育评价改革政策内容再生产程度仍存在明显差异,东部地区的内部差异性最大,中部地区的内部差异性最小。

综上所述,在对省级政府的政策内容再生产程度进行分析时,要充分考虑省级政府和省份的差异性,从内部和外部等多维视角分析政策内容再生产程度的影响因素。

通过对各省级政府对中央高等教育评价改革政策进行内容再生产的具体分析,研究发现,各省级政府都结合本省实际情况,围绕党委和政府教育工作评价、学校评价、教师评价、学生评价和用人评价等方面,对中央的高等教育评价改革政策做出了不同程度的细化与调适,而且政策内容再生产系数高的省份对中央政策的细化相对更加全面。

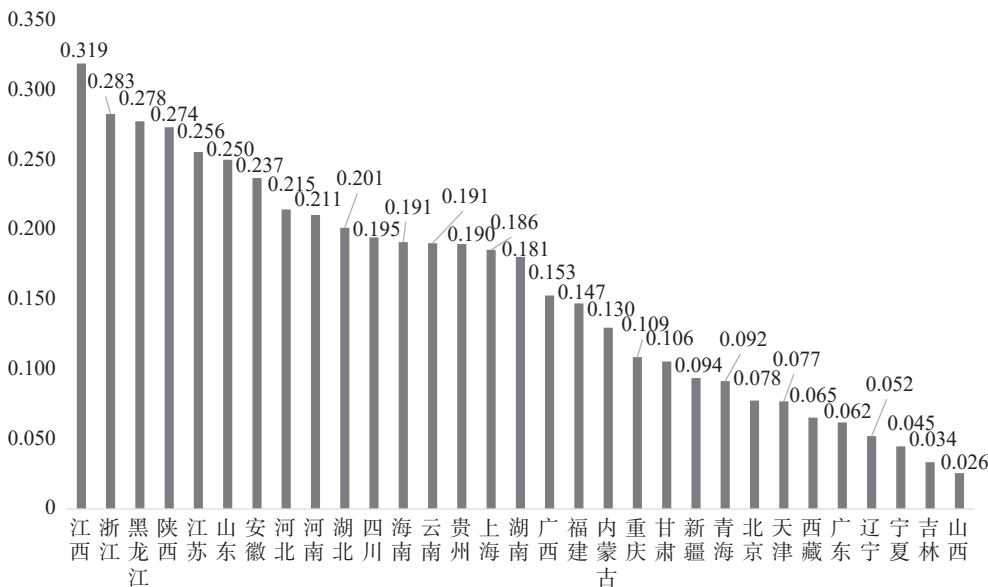


图 2 各省级政府高等教育评价改革政策内容再生产系数

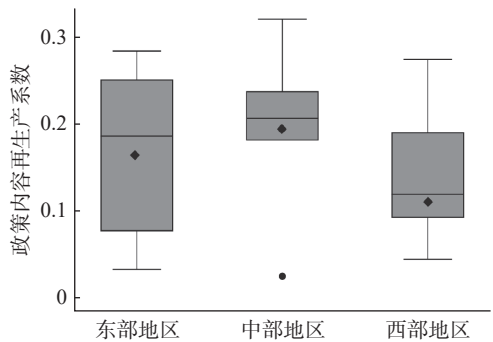


图 3 省级政府高等教育评价改革政策内容再生产系数的地区差异

省级政府进行政策内容再生产的意义在于将中央政策意图转化为符合地方实际的治理策略。因此,已有研究在对省级政府政策内容再生产程度进行探究时多隐含“政策内容再生产程度越高越好”

的假设,但并未对此进行检验。为了探究政策内容再生产程度是否越高越好,本研究进一步分析了省级政府高等教育评价改革政策内容再生产系数与高等教育综合发展实力之间的关系。为此,我们选取金平果科教评价网发布的《中国大学教育地区综合竞争力排行榜》2021—2024 年的均值作为衡量地区高等教育发展综合实力的指标^⑤(邵剑耀, 2024)。图 4 展示了二者之间的关系,省份的省级政府高等教育评价改革政策内容再生产程度与该省高等教育发展综合实力之间呈现显著的正相关关系($r=0.374$, $p<0.05$)。在一定程度上表明,省级政府高等教育评价改革政策内容再生产程度与政策实施的效能之间具有重要联系,结合散点分布情况可以发现,除了个别高等教育发展综合实力处于强弱两极的省份外,这种正相关关系在安徽、河北、河南、广西、重庆、甘肃、黑龙江等多数省份的表现均十分明显。这也从一个侧面表明,对于多数省份而言,省级政府结合地方实际对中央高等教育评价改革政策进行内容再生产是提升本省高等教育发展综合实力的有效举措之一。

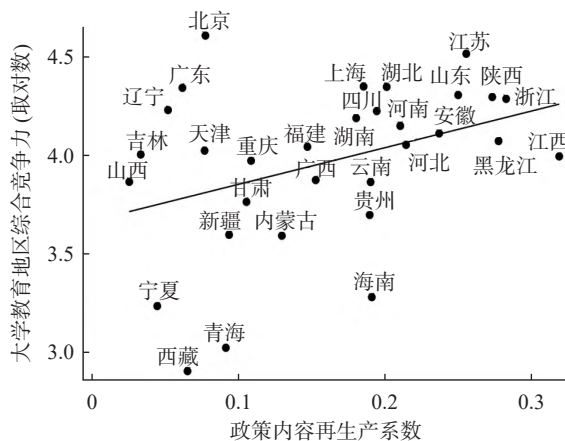


图 4 省级政府高等教育评价改革政策内容再生产程度与该省高等教育发展综合实力之间的关系

(二) 条件变量的单变量必要性分析

如果某一条件变量在结果发生时总是存在,则表明该条件变量是导致结果发生的必要条件(南锐等, 2025)。条件变量的单变量必要性分析是对所选取的条件变量中的单一变量对结果变量解释力的度量方法,即判断某一条件变量是否为结果变量的必要条件,这也是进行定性比较分析的先行步骤(王英伟, 2020)。一致性是衡量必要条件的重要指标,当一致性高于 0.900 时,则认为该变量是导致结果发生的必要条件(张明, 杜运周, 2019)。表 3 列出了分析结果,从中可见,7 个条件变量的一致性指标值均低于 0.900,说明单个条件变量对省级政府政策内容再生产程度的解释力较弱,多重因素并发联动才能有效解释省级政府政策内容再生产程度。

(三) 省级政府高等教育评价改革政策内容再生产程度的组态分析

运用 fsQCA 4.0 对数据进行处理,将案例数量的阈值设定为 1,一致性阈值设定为 0.8(程聪, 贾良定, 2016),通过标准分析对省级政府政策内容再生产程度高的条件组合进行识别,得到复杂解、中间解和简约解三种结果。纳入部分逻辑余项的中间解能够平衡理论与实际之间的关系,是定性比较分析结果汇报和条件组合路径诠释的首选方案(王法硕, 张嘉玲, 2023)。因此,本研究以中间解结果来呈现条件变量的组合,辅以简约解识别核心条件,以便说明各条件变量对省级政府政策内容再生产的相对重要性(Fiss, 2011)。表 4 列出了解释省级政府政策内容再生产程度高或不高的不同组态分析结果。

1. 政策内容再生产程度高的组态分析

由表 4 可见,应用 fsQCA 方法有效识别出省级政府高等教育评价改革政策内容再生产程度高的五种条件组态。省级政府政策内容再生产程度高的五种组态的一致性以及整体一致性均大于 0.800,表明所有组态均是高政策内容再生产程度的充分条件(石声萍等, 2020);整体解的覆盖度为 0.626,说明

五种条件组态可以解释 62.6% 的省级政府政策内容再生产程度高的案例, 组态覆盖的案例见表 4。

表 3 条件变量的单变量必要性分析结果

条件变量	政策内容再生产程度高	政策内容再生产程度不高
省级政府注意力配置	0.603	0.526
~省级政府注意力配置	0.705	0.729
省级教育财政投入	0.823	0.521
~省级教育财政投入	0.515	0.759
省级经济发展水平	0.647	0.608
~省级经济发展水平	0.653	0.642
省高等教育发展综合实力	0.782	0.532
~省高等教育发展综合实力	0.501	0.703
省际竞争	0.613	0.568
~省际竞争	0.657	0.656
中央政策信号	0.566	0.703
~中央政策信号	0.709	0.525
省级政府政策采纳速度	0.643	0.545
~省级政府政策采纳速度	0.618	0.670

注: 条件变量前没有“~”的表示该条件变量处于高值状态, 有“~”表示该条件变量不处于高值状态。“高”和“不高”的划分标准是变量校准后生成的模糊隶属分数。

表 4 省级政府高等教育评价改革政策内容再生产程度的组态分析

政策条件及其组态	政策内容再生产程度高					政策内容再生产程度不高		
	压力 传导型	横向竞争型		自主 优化型		资源缺乏型	压力异化型	
	G1	G2	G3	G4	G5	N1	N2	N3
省级政府注意力配置	▲	△	△	●	—	⊗	●	⊗
省级政府教育财政投入	⊗	▲	▲	▲	▲	△	△	△
省级经济发展水平	△	—	⊗	—	⊗	⊗	▲	▲
省高等教育发展综合实力	⊗	●	—	▲	▲	△	△	△
省际竞争	●	●	●	△	△	△	—	●
中央政策信号	●	△	△	△	△	—	⊗	●
省级政府政策采纳速度	●	●	●	⊗	⊗	△	—	●
一致性	0.859	0.834	0.887	0.853	0.885	0.867	0.937	0.956
原始覆盖度	0.226	0.346	0.314	0.346	0.325	0.374	0.218	0.203
唯一覆盖度	0.073	0.027	0.008	0.073	0.031	0.202	0.015	0.024
组态覆盖案例	海南	湖北 山东	江西 河南	江苏 安徽	河北 四川	青海 宁夏 山西 广西 甘肃	重庆 天津 福建	内蒙古
整体一致性	0.851					0.884		
整体覆盖度	0.626					0.500		

注: 1.核心条件是指同时存在于“简约解”和“中间解”的条件变量, ▲表示核心条件存在, △核心条件缺席。边缘条件是指仅出现在“中间解”的条件变量, ●表示边缘条件存在, ⊗边缘条件缺席, “—”表示该条件可有可无。2.原始覆盖度指的是给定组态覆盖结果案例的比例。3.唯一覆盖度是单个组态解释结果的程度。

根据分析框架以及核心条件分布的情况, 本研究将这五种条件组态归纳为如下三种类型: 压力传

导型、横向竞争型和自主优化型。在此基础上,结合2024年中央教育工作领导小组秘书组对部分省份教育评价改革落实情况的督察反馈意见,对这三种类型进行了分析。

(1)压力传导型(G1)。该类型呈现出外部压力驱动的特点,以中央政策信号、省际竞争和省级政府注意力配置作为进行较程度高等教育评价政策内容再生产的基础条件。具体表现为当省级政府接收到中央向其传递的政策信号并看到其他省级政府纷纷对此做出回应时,纵向压力和横向竞争成为驱动省级政府进行高等教育评价改革政策内容再生产的外部动力。省级政府会通过迅速调整注意力配置来进行较程度高等教育评价改革政策内容再生产,但呈现出政策采纳速度缓慢的特征。海南是这一组态的代表省份。海南省政府高度重视高等教育评价改革工作,在推动各级各类教育教学工作成效评价中强调“五育并举”、探索具有海南“特色印记”的育人成效指标体系等方面做了不少工作,但由外部压力驱动的政策内容再生产也表现出教育评价改革的某些做法仍带有明显的功利性,教育评价改革的科学性有待进一步提升。

(2)横向竞争型。这一类型的省级政府呈现出横向竞争的特点,以省际竞争作为推动高等教育评价改革政策内容再生产的外部压力。这一类型包括如下两种条件组态:

第一种条件组态(G2)的特点是:在省高等教育发展综合实力强的条件下,当省级政府面临横向竞争压力时,其会加大教育财政投入,自发加入高等教育评价改革政策内容再生产的横向竞争队列,积极推进高等教育评价改革政策内容的再生产,但政策采纳速度相对较慢。湖北、山东是这一组态的代表省份。例如,湖北在推进《总体方案》部署的各项改革任务时,结合地方高等教育发展实际,从干部履职能力提升、科技成果转让改革、学生综合素质评价等多个方面积极推进高等教育评价改革。

第二种条件组态(G3)的特点是:经济发展水平较低的一些省级政府同样会根据省际横向竞争压力调整教育财政投入,为高等教育评价改革政策内容再生产提供基本经济保障,但在政策实践中,其教育财政投入力度因经济发展水平低而相对更低,加上省高等教育发展综合实力相对较弱,使得该组态下的省级政府进行政策内容再生产时的速度相对更慢。采用这种条件组态的代表省份是江西和河南。例如,江西将推进教育评价改革情况纳入教育督导评价、在省属高校建立高层次人才专家举荐制度以及将美育工作及其效果作为高校办学评价的重要指标,纳入高校本科教学工作评估指标体系和“双一流”建设成效评价等。

由G1、G2和G3条件组态可以发现,省级政府的注意力配置以及高等教育财政投入可以有效消解外部压力,但这种由外部压力驱动的政策内容再生产过程常常比较缓慢,而且以纵向压力消解和横向竞争为动机的政策内容再生产虽然程度高,但存在的问题也比较突出。

(3)自主优化型。这一类型的省级政府呈现出即使没有外部压力也能自发快速的完成高等教育评价改革政策内容再生产的特点。这一类型包括如下两种条件组态:

第一种条件组态(G4)的特点是:省高等教育发展综合实力强劲,省级政府在高等教育领域配置的注意力强度较高,且具有高水平的教育财政投入,即使没有外部压力,省级政府也能够自发地进行高等教育评价改革政策内容的再生产,且采纳中央政策的速度较快。江苏和安徽是这一组态覆盖的省份。例如,江苏省政府高度重视高等教育评价改革工作,围绕“高水平大学建设工程”相关工作多次做出部署,特别是江苏在2018年便提出了高水平大学建设绩效评价指标体系,2019年进行了修订与完善,该指标体系与《总体方案》提出的改革方向相一致(余达淮,邹阳,2021)。可见,由政府自驱力主导的政策内容再生产表现出速度快且质量高的特征。

第二种条件组态(G5)的特点是:省高等教育发展综合实力强,具备开展高等教育评价改革的良好基础,但经济发展水平相对落后,省级政府通过教育财政投入为高等教育评价改革政策内容再生产提供经济支持。在这种情况下,省高等教育发展综合实力与教育财政投入力度的契合也能促使省级政府快速完成高等教育评价改革政策内容的再生产,但在具体政策实践中,政策内容再生产过程和程度明

显低于G4组态。河北和四川是这一组态的代表省份。例如,河北在中央《总体方案》发布后,积极推进各项改革任务落实落地,坚持清单式管理、台账式推进,并成立河北省教育质量评估监测中心,为工作推进提供专业人才支持。

对上述两种条件组态的比较可以发现,省高等教育发展综合实力强并具备相应规模的高等教育财政投入是省级政府自发完成高等教育评价改革政策内容再生产的重要前提。此外,若省级政府在高等教育评价改革政策方面的注意力配置强,也会一定程度上提高政策内容再生产的速度和程度。

通过对上述五种条件组态的代表性案例分析发现,G1、G3和G5条件组态对应经济相对欠发达省份的高等教育评价改革政策内容再生产情况,G2和G4条件组态对应经济相对发达省份的高等教育评价改革政策内容再生产情况。对于经济欠发达省份,省级政府的注意力配置或教育财政投入是进行高等教育评价改革政策内容再生产的基础条件,二者均能够有效回应外部压力,从而实现高程度的高等教育评价改革政策内容再生产。对经济发达省份而言,省级政府的自发性和应对横向竞争压力均会触发较高程度的政策内容再生产,但自主优化型路径的一致性和唯一覆盖度相对较高,表明该路径对结果的解释力度更强。此外,省级政府政策采纳速度与政策内容再生产程度二者之间并非简单的线性关系,说明省级政府政策采纳速度并不一定是越快越好。从侧面表明了考察政策传递过程中省级政府政策内容再生产的必要性,因为单纯考察省级政府政策采纳速度难以反映其政策采纳过程的全貌。

2.政策内容再生产程度不高的组态分析

为进一步探索省级政府高等教育评价改革政策内容再生产的驱动机制,本研究也分析了政策内容再生产程度不高的组态(表4)。由表4可知,可以解释政策内容再生产程度不高的组态有三种,它们的整体一致性为0.884,整体覆盖度为0.500,表明所有组态是政策内容再生产程度不高的充分条件,能够解释50.00%的政策内容再生产程度不高案例。根据核心条件的存在情况,可以将三种组态划分为两种类型:资源缺乏型与压力异化型。

(1)资源缺乏型。该类型表明当不存在外部压力时,资源缺乏是省级政府政策内容再生产程度低的重要影响因素。这一类型包括如下两种组态:

第一种组态(N1)的特点是:省高等教育发展综合实力弱,省级政府对高等教育评价改革工作的注意力配置、教育财政投入以及经济发展水平的缺失使得省级政府只能充当中央文件的“二传手”,快速采纳并向下传递中央政策。青海、宁夏、山西、广西和甘肃是这一组态覆盖的省份。这些省份在推进高等教育评价改革时,存在评价方式创新性不足、专业化队伍建设亟待提高以及改革力度和时效性有待加强等问题,需要省级政府进一步重视相关工作并抓好机制建设。

第二种组态(N2)的特点是:虽然省级政府相对重视高等教育评价改革工作,但省级高等教育发展基础较弱,高等教育财政投入不足。加上不存在外部压力,即使经济发展水平较高,省级政府最终仍呈现出高等教育评价改革政策内容再生产程度低的结果。重庆、天津、福建是这一组态的代表省份。例如,重庆在中央《总体方案》出台后,出台了42个市级配套文件进一步完善教育评价制度体系,但在高等教育领域基础设施建设及科创平台建设方面仍需加强。

虽然两种组态都会导致政策内容再生产程度低的结果,但通过对比代表省份的政策内容再生产系数可以发现,省级政府的注意力配置和经济发展水平对政策内容再生产程度的提升具有一定影响。此外,N1组态的原始覆盖度更高,表明因政府内驱力不足和发展基础薄弱导致的“高速度-低程度”的政策内容再生产模式更普遍。

(2)压力异化型(N3)。该类型表明如果省级政府面临较大的外部压力,且省级政府进行政策内容再生产的内生动力不足时,外部压力会使得政策内容再生产发生异化,最终导致“低速度-低程度”的结果。内蒙古是这一组态的代表省份。内蒙古自治区政府在《总体方案》出台后,发布了一系列涉及高等教育评价改革的政策文件,但存在评价标准细化不足、高校教师科研评价工作中仍存在“重数量轻质

量”等问题。

尽管政策内容再生产程度高和不高的组态路径具有非对称性特征,但对比高与不高的组态分析结果可以发现:

第一,省级政府的高等教育财政投入是促进其进行高等教育评价改革政策内容再生产的经济基础。具体表现为,在政策内容再生产程度高的五种组态中,省级政府教育财政投入作为四种组态的核心条件存在,而在政策内容再生产程度不高的组态中其均作为核心条件缺失。此外,通过对比 G1 与 N3 组态、G2、G3 与 N3 组态可以发现,外部压力驱动的政策内容再生产表现出速度缓慢的特征,但外部压力的存在会促使省级政府调整其注意力配置或提供相应的教育财政投入,从而提高政策内容再生产程度。这一结果说明省级政府在面对外部压力时,若能够以调整注意力配置或财政投入等方式积极回应外部压力,也可提高政策内容再生产程度。这也从一个侧面说明了财政投入对高等教育评价改革政策内容再生产的作用。

第二,省高等教育发展综合实力是省级政府进行高等教育评价改革政策内容再生产的现实基础。具体表现为,在政策内容再生产程度高的五种组态中,省高等教育综合实力作为其中两种组态的核心条件存在,在政策内容再生产程度不高的三种组态中,其均作为核心条件缺失。此外,通过对比自主优化型和资源缺乏型组态可以发现,在不存在外部压力的情况下,省高等教育发展综合实力与教育财政投入的契合是推动省级政府进行高质量政策内容再生产的核心驱动因素。

五、结论与讨论

(一) 研究结论

高等教育评价改革是推动高等教育高质量发展的重要举措。2020 年以来,中央政府逐步完善了我国高等教育评价改革政策的顶层设计。但中央政策的贯彻落实要求省级政府根据地方实际因地制宜地制定相应方案,才能使中央政策有效落地。省级政府对高等教育评价改革政策内容的再生产反映了其在政策贯彻落实过程中是否结合地方实际对中央政策进行了细化和更新,这对中央政策的实施效果具有重要影响。已有研究虽然注意到了在高等教育评价改革政策传递过程中出现的省际趋同现象,但囿于测量方法的限制并未对其进行深入的探讨和分析。有鉴于此,本研究基于机器学习方法,运用 Doc2vec 模型对省级政府高等教育评价改革政策内容再生产进行了测量,并运用模糊集定性比较分析方法对高等教育评价改革政策内容再生产的影响机制进行了条件组态分析。研究的主要结论有:

1.省级政府对高等教育评价改革政策内容再生产程度偏低。我国 31 个省级政府的高等教育评价改革政策内容再生产系数均值为 0.159,多数省份发布的政策与中央政策存在内容的高度相似性,就整体而言,政策内容再生产程度偏低。研究发现,省级政府对高等教育评价改革政策内容再生产的程度与该省高等教育发展综合实力存在正相关关系。这意味着提高省级政府高等教育评价改革政策内容再生产程度,有助于促进其高等教育综合实力的提升。

2.省级政府高等教育评价改革政策内容再生产受多重因素的影响。基于政策创新扩散理论,本研究从内外部两个方面对影响省级政府高等教育评价改革政策内容再生产的因素进行了分析,并将这些因素分为省级政府注意力配置、省级政府教育财政投入、省级经济发展水平、省高等教育发展综合实力、省际竞争、中央政策信号与省级政府政策采纳速度 7 个方面。单变量必要性分析结果表明,这 7 个因素中任何一个因素均不构成省级政府高等教育评价改革政策内容再生产程度的必要条件,这些因素并发联动才能对其进行有效解释。

3.不同省级政府实现较程度的高等教育评价改革政策内容再生产所采用的路径不同。组态分析结果表明,省级政府实现较程度的高等教育评价改革政策内容再生产共采用了五种不同的条件组态,这五种条件组态又可以进一步概括为三种类型,即压力传导型、横向竞争型和自主优化型。由于

各省的实际情况不同,它们进行高等教育评价改革政策内容再生产所采用的路径不同。例如,经济欠发达地区的省级政府进行高等教育评价改革政策内容再生产的动力主要源于外部压力,但外部压力会直接影响省级政府进行政策内容再生产的动机,可能会带来政策内容再生产程度高、速度慢且质量低的问题;而经济发达地区的省级政府进行高等教育评价改革政策内容再生产的内驱力更大,进行政策内容再生产的积极性更高。

4.高等教育评价改革政策内容再生产程度不高的类型可概括为资源缺乏型和压力异化型。高等教育评价改革政策内容再生产程度不高的组态有三种,这三种组态可分为两种类型:资源缺乏型与压力异化型。通过比较高等教育评价改革政策内容再生产程度高与不高的组态可以发现,省级政府教育财政投入是其进行政策内容再生产的经济基础,而高等教育发展综合实力是其进行政策内容再生产的现实基础。一方面,省级政府高等教育财政投入与其注意力配置具有替代效应,二者均能积极回应外部压力并实现高程度的高等教育评价改革政策内容再生产;另一方面,其与省高等教育发展综合实力具有协同效应,二者的配合能够强化省级政府政策内容再生产的内驱力,进而促使省级政府进行高质量的政策内容再生产。

(二) 政策启示

本研究结果对于进一步推动高等教育评价改革政策落地以及完善高等教育评价改革政策体系具有一定的启示:

1.省级政府要因地制宜,进一步细化、深化中央高等教育评价改革的相关政策。本研究结果表明,省级政府结合本地区实际,对中央有关高等教育评价改革政策进行细化、深化的努力不够,这必然会在一定程度上影响中央政策的执行及其效果。为此,省级政府要根据中央的相关政策,从本地区高等教育评价改革面临的实际问题出发,因地制宜制订更加具体、更有针对性和地方特色的贯彻落实方案和措施,切实推进高等教育评价改革政策落地,从而充分激发政策效能,提升高等教育发展综合实力。

2.省级政府应积极培育进行政策内容再生产的内生动力。研究结论表明,省级政府基于自身特征与辖区情境进行的政策内容再生产表现出速度快且程度高的特征,而外部压力虽然也能促使省级政府进行高程度的政策内容再生产,但可能导致政策内容再生产存在高程度与低质量并存的问题,且可能会因省级政府的消极回应而发生异化,并导致政策内容再生产程度低的后果。因此,各省级政府应以促进高等教育内涵式发展和切实提升高等教育质量为目标,积极培育政策内容再生产内生动力。具体而言,一方面要深入树立“以人民为中心”的发展理念,以提高教育财政投入等现实举措回应本地区对高等教育评价改革的期盼和诉求;另一方面要根据本地区高等教育发展的现实情境合理设置政策优先级,优化省级政府的注意力配置和强化资源投入。

3.中央政府要进一步加强督导检查及奖罚力度。在当前的治理体制下,省级政府基于合法性压力和横向竞争压力而采纳中央政策,但将采纳行为扩展至政策内容再生产时,合法性压力的影响会减弱,这就要求中央政府要针对省级政府的政策内容再生产设置督导检查与考核激励机制,对高等教育评价改革政策贯彻落实有力、效果好的省级政府要给予表彰奖励,对贯彻落实不力、效果差的省级政府要采取问责措施。

本研究也存在一定局限性:一方面,由于本研究仅聚焦于高等教育评价改革政策,该分析框架的普适性还需引入更多教育评价改革政策加以检验。另一方面,本研究仅根据省级政府的相关政策文件进行研究,难以捕捉省级政府在高等教育评价改革实践中的具体行动表现。因此,对省级政府高等教育评价改革政策内容再生产的研究,未来还需要结合政策实践进一步展开,以期更全面地还原省级政府高等教育评价改革政策内容再生产的过程和实现机制。

(陆根书工作邮箱: gslu@xjtu.edu.cn; 董宇婧为本文通信作者: dongyujing96@163.com)

参考文献

- 陈刚,林志义,林向阳.(2023).我国体育政策政府间传递与再生产机制研究——针对国家和省级政府体育政策的文本考察. *西安体育学院学报*, 40(02), 167—177.
- 陈洁.(2023).基于 Doc2Vec 增强特征的文本文主题聚类研究. *计算机科学*, 50(S1), 221—226.
- 程聪,贾良定.(2016).我国企业跨国并购驱动机制研究——基于清晰集的定性比较分析. *南开管理评论*, 19(06), 113—121.
- 邓雪琳.(2015).改革开放以来中国政府职能转变的测量——基于国务院政府工作报告(1978-2015)的文本分析. *中国行政管理*, (08), 30—36.
- 杜运周,刘秋辰,陈凯薇,肖仁桥,李姗姗.(2022).营商环境生态、全要素生产率与城市高质量发展的多元模式——基于复杂系统观的组态分析. *管理世界*, 38(09), 127—145.
- 郭伟.(2021). *中国可持续旅游政策创新扩散的影响因素及相互作用研究*. 广州:华南理工大学博士学位论文.
- 黄荣贵.(2017).网络场域、文化认同与劳工关注社群 基于话题模型与社群侦测的大数据分析. *社会*, 3(02), 26—50.
- 刘振天.(2024).高等教育评价改革的主要成效、基本经验与完善之策. *中国高等教育*, (Z2), 21—26.
- 刘河庆,梁玉成.(2021).政策内容再生产的影响机制——基于涉农政策文本的研究. *社会学研究*, 36(01), 115—136+228-229.
- 刘河庆.(2020).文件治理中的政策采纳及其影响因素研究 基于国家和省级政府政策文本(2008—2018)数据. *社会*, 40(04), 217—240.
- 林雪霏.(2015).政府间组织学习与政策再生产:政策扩散的微观机制——以“城市网格化管理”政策为例. *公共管理学报*, 12(01), 11—23+153-154.
- 凌磊.(2018).我国自主招生考试政策影响因素及实现路径——基于政策创新扩散理论的实证分析. *复旦教育论坛*, 16(03), 51—57.
- 李瑞昌.(2012).中国公共政策实施中的“政策空传”现象研究. *公共行政评论*, 5(03), 59—85+180.
- 李欢欢,顾丽梅.(2020).垃圾分类政策试点扩散的逻辑分析——基于中国 235 个城市的实证研究. *中国行政管理*, (08), 81—87.
- 南锐,肖叶静,曹丹婷.(2025).数据赋能何以影响政府治理效率?——基于中国 31 个省份的 fsQCA 分析. *电子政务*, (05), 89—100.
- 秦一帆.(2024).省域推进新时代教育评价改革的困境与出路——基于 20 个省(市)实施方案的政策分析. *教育评论*, (04), 70—76.
- 仇叶.(2021).县级政策转换与有效治理——对中国公共政策过程的反思. *经济社会体制比较*, (03), 99—108.
- 阮光册,夏磊.(2019).基于 Doc2Vec 的期刊论文热点选题识别. *情报理论与实践*, 42(04), 107—111+106.
- 任锋,朱旭峰.(2010).转型期中国公共意识形态政策的议程设置——以高校思政教育十六号文件为例. *开放时代*, (06), 68—82.
- 邵剑耀.(2024).软实力助力抑或硬实力支撑:省域高等教育发展路径研究. *高校教育管理*, 18(04), 40—50.
- 石声萍,何新月,杨刚,洪静,王双龙.(2020).政策与文化因素组态效应对企业创新的影响. *外国经济与管理*, 42(12), 89—103.
- 王法硕,张嘉玲.(2023).“互联网+政务服务”政策再生产及其影响因素——基于机器学习与定性比较分析的实证研究. *公共管理与政策评论*, 12(04), 47—60.
- 王英伟.(2020).权威援引、资源整合与外压中和:邻避抗争治理中政策工具的选择逻辑——基于(fsQCA)模糊集定性比较分析. *公共管理学报*, 17(02), 27—39+166.
- 王洪涛,陈洪侠.(2017).我国智慧城市创新扩散演进机理及启示——基于 38 个城市的事件史分析. *科技进步与对策*, 34(03), 44—48.
- 吴柯豫,魏署光.(2025).职业本科教育试点政策创新扩散的逻辑动因探究——基于全国 31 个省域数据的事件史分析. *大学教育科学*, (02), 99—111.
- 吴峰,石益静.(2024).教育信息化政策创新扩散的时空特征与影响因素. *苏州大学学报(教育科学版)*, 12(02), 82—92.
- 薛增鑫,满小欧.(2024).层级治理体系下的政策再生产、注意力配置与政策效应——基于双重机器学习模型的数字乡村政策执行研究. *华中科技大学学报(社会科学版)*, 38(03), 117—128.
- 薛阳,李曼竹,高云红,冯银虎.(2023).中国碳排放权交易制度的扩散机制研究——基于宏观—微观互动逻辑的事件史分析. *城市问题*, (07), 43—52.
- 阳荣威,刘伟豪.(2023).新时代教育评价改革政策执行的阻滞因素与纾解路径. *大学教育科学*, (02), 51—60.
- 易兰丽,范梓腾.(2022).层级治理体系下的政策注意力识别偏好与政策采纳——以省级“互联网+政务服务”平台建设为例. *公共管理学报*, 19(01), 40—51+167.
- 易凌云.(2021).“五唯”问题:实质与出路. *教育研究*, 42(01), 4—14.
- 余达淮,邹阳.(2021).推进教育评价改革,促进高水平大学建设——以江苏高水平大学建设绩效评价改革为例. *高校教育管理*, 14(02), 37—45.
- 杨宇婷,王名扬,田宪允,李鹏宇.(2016).基于文档分布式表达的新浪微博情感分类研究. *情报杂志*, 35(02), 151—156.
- 张晓君,王伟桥.(2024).地方政府应急管理五年规划内容再生产——影响因素与实证分析. *软科学*, 38(10), 47—54.
- 张辉蓉,盛雅琦.(2022).新时代教育评价改革的制度化困境及应对策略——新制度主义的视角. *中国电化教育*, (07), 80—86.
- 张海柱,林华旌.(2022).政策扩散中“政策再创新”的生成路径与内在逻辑——基于 16 个案例的定性比较分析. *公共管理学报*, 19(01), 27—39+166.

- 张明,杜运周.(2019).组织与管理研究中QCA方法的应用:定位、策略和方向.《管理学报》,16(09),1312—1323.
- 周光礼.(2023).建构中国特色高等教育评价体系.《教育研究》,44(08),4—14.
- 周洪宇.(2020).深化教育评价改革 加快推进教育现代化——《深化新时代教育评价改革总体方案》解读.《中国考试》,(11),1—8.
- Augier, M. (2002). Administrative Behavior: A Study of Decision - Making Processes in Administrative Organizations. *The Economic Journal*, 112(480), 386—388.
- Berry, F. S., & Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. *American political science review*, 84(2), 395—415.
- Berry, F. S., & Berry, W. D. (2018). Innovation and diffusion models in policy research. *Theories of the policy process*, 253—297.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993—1022.
- Chen, J., & Huang, C. (2021). Policy reinvention and diffusion: Evidence from Chinese provincial governments. *Journal of Chinese Political Science*, 26(4), 723—741.
- Crilly, D., Zollo, M., & Hansen, M. T. (2012). Faking it or muddling through? Understanding decoupling in response to stakeholder pressures. *Academy of Management Journal*, 55(6), 1429—1448.
- Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of management journal*, 54(2), 393—420.
- Hinkle, R. K. (2015). Into the words: Using statutory text to explore the impact of federal courts on state policy diffusion. *American Journal of Political Science*, 59(4), 1002—1021.
- Jansa, J. M., Hansen, E. R., & Gray, V. H. (2019). Copy and paste lawmaking: Legislative professionalism and policy reinvention in the states. *American Politics Research*, 47(4), 739—767.
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning*. PMLR.
- Mallinson, D. J. (2016). Building a better speed trap: Measuring policy adoption speed in the American states. *State Politics & Policy Quarterly*, 16(1), 98—120.
- Ragin, C. C. (2009). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.
- Shipan, C. R., & Volden, C. (2006). Bottom - up federalism: The diffusion of antismoking policies from US cities to states. *American journal of political science*, 50(4), 825—843.
- Volden, C., Ting, M. M., & Carpenter, D. P. (2008). A formal model of learning and policy diffusion. *American Political Science Review*, 102(3), 319—332.
- Walker, J. L. (1969). The diffusion of innovations among the American states. *American political science review*, 63(3), 880—899.
- Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and performance*, 18(Chapter 3), 73—103.

注 释:

① 因果关系的非对称性是定性比较分析 (Qualitative Comparative Analysis, QCA) 方法的核心特征, 伯努瓦·里豪克斯 (Benoît Rihoux) 和查尔斯 C·拉金 (Charles C. Ragin) 在《QCA 设计原理与应用: 超越定性定量研究的新方法》这本书中指出, 所谓因果关系的非对称性, 是指某个结果的出现与否可能需要不同的“原因组合”来分别解释。

② 《总体方案》是中央指导新时代教育评价改革的总纲领, 从战略性任务的高度部署了改革党委和政府教育工作评价、学校评价、教师评价、学生评价和用人单位评价五项重点任务, 但关于改革党委和政府教育工作评价的任务涉及面广, 尚未单独出台针对高等教育领域的文件, 因此选取涵盖全部教育领域的政策文件, 而国家对于高等教育评价改革的工作部署是在教育评价改革大背景下展开的, 所以将《总体方案》也纳入分析范围。此外, 考虑到本科教育和研究生教育评价工作存在较大异质性, 此处的高等教育仅指本科高等教育阶段。

③ 政府工作报告主要分为上一年工作回顾、新一年目标任务和工作内容等部分。鉴于省级政府注意力配置意在测量其对新一年各项政策议题和工作事务的重视情况, 我们在实际处理时只保留了新一年目标任务和工作内容等部分, 将上一年工作回顾部分进行了剔除。

④ LDA 主题模型是典型的无监督模型, 其核心目标是从文本数据中发现隐含的主题。具体而言, 主题由一组词汇的统计分布所定义, 模型结果既可以估计出每个主题所对应的关键词及出现概率, 也可以估计出每一个文本所对应的各主题的出现概率。

⑤ 为了平滑变量之间的数量级差距、避免极端值的影响, 这里对大学教育地区竞争力指标进行了对数化处理。

(责任编辑 王 森)

Policy Reproduction of Provincial Governments' Higher Education Evaluation Reforms and the Influence Mechanisms: An Empirical Research Based on Machine Learning and Qualitative Comparative Analysis

Lu Genshu^{1,2} Dong Yujing¹ Liu Yangyun³

(1. School of Humanities and Social Sciences, Xi'an Jiaotong University, Xi'an 710049, China;

2. Western China Higher Education Evaluation Center, Xi'an Jiaotong University, Xi'an 710049, China;

3. Professional Committee on Educational Evaluation, Chinese Society of Educational Development Strategy, Beijing 100084, China)

Abstract: Higher education evaluation reform serves as a critical component in enhancing China's comprehensive higher education capacity and establishing the nation as a global higher education powerhouse. Provincial governments' policy reproduction during the formulation of evaluation reform policies significantly impacts local practices. Based on the machine learning method, a Doc2vec model was constructed to conduct quantitative calculations on the reproduction of higher education evaluation reform policy content in China's provinces. Furthermore, Qualitative Comparative Analysis was employed to uncover the influence mechanisms of provincial governments' policy reproduction in higher education evaluation reform. The research finds that the degree of policy reproduction in provincial higher education evaluation reforms remains generally low. Five high-reproduction configurations were identified, which can be further categorized into three pathways: pressure-driven, horizontal competition-driven, and autonomous optimization. Three low-reproduction configurations were identified, which can be further generalized into two distinct pathway typologies: resource-deficient type and pressure-alienated type. A comparative analysis of pathways with high and low reproduction reveals that the provincial government's education fiscal input is the economic basis for the provincial government to promote the reproduction of policy content of higher education evaluation reform, and the comprehensive strength of provincial higher education is the realistic basis for the provincial government to promote the reproduction of policy content of higher education evaluation reform. There is a substitution effect between the provincial government's education fiscal input and the attention allocation of the provincial government, both of which can respond positively to external pressure. When there is no external pressure, they have a synergistic effect with the comprehensive strength of provincial higher education development, and their coordination can effectively enhance the degree of policy content reproduction by provincial governments. Therefore, provincial governments should stimulate internal motivation, rationally allocate education fiscal input, and implement central policies in a way that suits local conditions. The central government should further strengthen supervision, inspection, and reward and punishment measures to better promote the implementation of relevant policies.

Keywords: higher education evaluation reform policies; policy reproduction; machine learning; fsQCA