智能体德尔菲法驱动的高职专业质量评价标准重构

杨倬东
(巴中职业技术学院,四川巴中　636000)

[摘　要]针对某高职院校首轮专业质量评价中政策响应维度缺失、实践效能区分度失灵的问题,引入人工智能赋能标准修订。循"政策对标—问题诊断—AI赋能—标准重建"路径,以政策文本与23个专业数据双重审查既有标准,设计"8个AI专家角色+人类审核裁决"双层架构,借AI辅助德尔菲法三轮迭代并检验共识度与信度,进而将三阶框架重构为"定位—教学条件—人才培养—师资—产教融合"五维体系,规模类数据"监测不评价",并嵌入"监测—评估—整改—复评"PDCA闭环。研究提出"人机协作"标准制定与"数据驱动PDCA(D-PDCA)"治理双范式,可资同类院校借鉴。

[关键词]人工智能;专业质量评价;评价标准;德尔菲法;PDCA闭环
[中图分类号]G718.5;G40-058.1　　[文献标识码]A　　[文章编号]2096-711X(2026)20-0085-04
doi:10.3969/j.issn.2096-711X.2026.20.028　　[本刊网址]http://www.hbxb.net

专业是职业院校人才培养与教学工作的核心,其建设质量既牵系学生就业竞争力,又标定学校办学水平;而"有什么样的评价指挥棒,就有什么样的办学导向",评价标准一体两面,不可偏废。当下质量观由"小逻辑"转向"大逻辑","适应性"本就"起源于生态学",被提为"职业教育本质属性之一",要求人才培养于内适性与外适性间双向统一。此一转型获政策与技术双重加持:教育部等五部门印发的《"人工智能+教育"行动计划》提出"深入开展人工智能赋能教育评价……构建人机协同的教育新形态";《四川省教育数字化行动计划(2025—2027年)》要求"建立数据驱动的质量评价机制,加强教育质量监测评估体系建设";《职业院校人工智能应用指引(2026版)》明确"职业院校应将人工智能与智慧校园已有技术系统进行集成……提供智能化技术与工具的支持"。而智能模型的跃迁,使传统德尔菲法久困于成本高、周期长、权威与从众之扰,有望借"AI智能体专家角色"的低成本、可复现模拟而得纾解。本研究问题意识源于一次真实的评价实践:某高职院校试行的首轮标准覆盖23个专业,虽搭起骨架却两端暴露病灶——对照政策,课程与教材、经费保障、数字化与智能化等要素几近空白;揆诸数据,大量指标或全域获D等而沦为"失效指标",或全域获A等而呈"天花板效应"。正是"可评性"与"诊断性"间的裂隙,催生本研究循"政策对标—问题诊断—AI赋能—标准重建"之进路重构标准体系。

一、首轮专业质量评价标准的制定背景与基本架构

首轮标准的制定,根植于职业教育"诊断与改进"的制度土壤。在"管办评"分离下,质量保障体系"由原来的外部质量保障体系强势、内部质量保障体系弱势转变为内外并重的质量保障体系治理结构",专业作为人才培养的基层组织单位,其内部评价"能够为高职院校开展差异化专业建设提供依据",是诊改落地的重要抓手。该校因专业数量多、类别跨度大、缺乏统一可比的质量尺度,遂依托人才培养工作状态数据平台、教务系统与就业报告等,研制了覆盖输入、过程、产出全链条的首轮指标体系。

就架构而言,首轮采用"专业基础(输入)—教学运行(过程)—发展成效(产出)"的三阶框架,与学界广用的"投入—过程—产出"(I-P-O)模型相契合,体现了高职教育"由资源输入、教育教学实施、成果输出的循环往复的过程"这一规律。三阶之下分设二级指标与观测点,采用A、C、D三级判定,与同类研究"分等级确定达标标准""定性结合定量"一脉相承。然而,所沿用的传统人类专家咨询难脱权重"往往非常主观,由评价举办方根据重要性进行自行设定"的局限,致使指标的政策响应性与实证适配性未获充分校验。

二、首轮评价标准的问题审视——基于政策对标与实践反思的双重视角

对首轮标准的审视遵循"政策对标"与"实践反思"双重视角,一外一内,共同回答"缺什么、失什么"。

(一)政策对标分析:基于最新政策文本的指标覆盖度审查

《四川省高等职业学校办学能力评价实施方案(2025—2030年)》构建了"办学条件常态化监测、教学工作定期评估相结合的评价体系",其监测指标"包括4个一级指标、21个二级指标、46个监测点"。《高等学校人才培养质量保障体系建设指南》则要求"建立健全分层分类、清晰可测、公开透明、覆盖人才培养全要素全过程的质量标准体系……构建覆盖全要素全过程的智能监测与诊断系统"。对标可数出三类缺失:其一维度缺失,对信息化条件、经费保障等投入类指标未作覆盖,对"课程与教材"仅以"课程标准规范性"一带而过;其二机制缺失,未建"动态监测"机制与"评价—诊断—反馈—整改"闭环;其三将"学生规模"纳入评价,易致"以大为优"的逆向激励。

教育部《关于深化职业教育教学关键要素改革的意见》

收稿日期:2026-8-28
基金项目:本文系四川省职业技术教育学会一般课题"AI赋能职业院校专业质量评价标准体系构建与实践研究"阶段性研究成果(项目编号:Y251024)。
作者简介:杨倬东(1988—),男,四川巴中人,巴中职业技术学院副教授,主要从事教育教学评价、教育教学管理数智化转型研究。

以专业、课程、教材、教师、实习实训五大关键要素改革为主线,对标更显差距:政策要求"利用大数据和人工智能精准预测关键领域人才培养供需情况",首轮仅设"就业率"单一产出指标;政策要求"绘制课程能力图谱",首轮仅考察"课程标准规范性";政策推行"三主编"机制与活页式、工作手册式教材,首轮未设此项;政策要求"研究制定教师能力清单",首轮仅以"双师比例"作粗判;政策强调"产教融合实习实训基地",首轮虽设"校企合作"观测点但隐于"教学运行"。凡此表明首轮标准对最新政策的响应存在系统性滞后。

(二)实践反思分析:基于首轮评价实施数据的指标效能审查

若说政策对标照见"缺什么",实施数据则照见"失什么"。就区分度看,一是全域D等、区分度为零的"失效指标",如"双元育人模式""专业带头人能力""名师、工作室或名团队"等全部或近乎全部获D等;二是"天花板效应"下的低区分度指标,如"人才培养方案规范性、科学性""课程开课执行率"几乎全A,仅验文件之完备而不见质量之参差;三是真正具有区分效能者,如生师比、双师型占比、校内外实践条件差异显著,共性在于有可量化阈限、数据可追溯。就维度均衡看,指标类型严重失衡,尤以"社会评价"维度数据缺失致其虚置为甚,恰呼应徐春艳所指外部评价的"信息不对称"困境,"企业自身的技术革新、岗位需求变化等信息也不能及时反馈给高职院校"。凡此汇成"可评性"与"诊断性"的根本对立,由此可得五项启示:指标须穿透"文件层"进入"实施层"、参照同类院校实证分布、体现政策新要求、证据均衡、数据采集内嵌于日常教学管理。

三、AI赋能专业质量评价标准修订的策略与路径

针对上述病灶,研究提出以人工智能赋能修订的核心策略、规范化流程与一致性检验适配。

(一)核心策略:构建AI智能体专家团队的人机协作修订模式

标准修订本质上是专家智慧集成的工作,其共同逻辑是"聚合多元主体意见、消除偏见、形成共识"。传统人类专家咨询受限于时间成本,易受权威效应与从众心理影响,迭代周期长且难以复现。人工智能大模型为破解此困局提供新路径:通过设定不同专业背景、价值立场、角色定位的AI智能体专家,可低成本、可复现地模拟多元主体参与。《职业院校人工智能应用指引(2026版)》指出:"'以人为本'的人工智能意识,强调学生在人工智能应用中的主体意识与技术态度,是学生形成正确理念的重要基础。"据此,本研究构建"8个AI智能体专家角色+人类审核裁决"的双层架构:AI层负责多视角方案生成与迭代收敛,人类层负责最终裁决、伦理把关与可操作性校准。

每个AI专家角色遵循核心知识域声明、思维倾向描述、评分偏好公开与潜在偏见警示的"四要素"原则,并在独立上下文中作答以保障评判互不干扰,正是对"评价主体应扩展到包括政府、学校、行业企业、第三方机构、学生和教师等多元主体"之"大逻辑"的技术化实现。人机协作坚守"人类主导、AI辅助":人类研究者承担指标池初建、角色设计、轮次调度与最终审核的职责,AI协调者负责匿名化与统计量计算,AI专家团队进行独立评分与群体参照下的修订,最终价值判断权始终握在人类手中。

(二)AI赋能路径与人类专家路径的对比验证

为检验AI辅助德尔菲法的有效性与局限性,本研究将其成果与同期基于传统人类专家咨询的《专业质量监测评估办法修订研究报告》(简称"HI路径")对比。HI路径遵循"政策对标—数据反馈—利益相关者意见征求"三阶推进,面向7类利益相关者征求意见,收到47条建议、采纳35条(采纳率74.5%)。两条路径的最终成果殊途同归:关键改进高度趋同——均将"产教融合"从原"教学运行"维度独立出来、均将"学生规模"调整为"监测不评价"、均将"双师比例"阈值由40%提至50%、均增设"监测—评估—整改—复评"闭环机制。这种"双盲"条件下的成果趋同,从实证层面验证了AI辅助德尔菲法作为专家智慧集成工具的有效性。

然对比亦照见AI路径的短板:一是利益相关者诉求捕捉不足,未能覆盖"基层教师的操作焦虑"这一隐性维度;二是措辞精细化的依赖,HI路径在反馈后将"考核优秀"改为"考核合格"以照顾基层,AI路径在此类"温度调节"上仍显机械;三是价值判断的依赖,AI可生成方案但无法替代人文关怀的终极追问。故"AI+HI"范式的合理性不在于替代,而在于AI以规模化模拟降低初期成本,人类专家在伦理把关、价值裁决环节发挥"最后一公里"的决定作用。

(三)规范化工作流程:AI辅助德尔菲法的三轮迭代设计

德尔菲法的精髓在于匿名、反馈与多轮收敛。刘静给出的规范步骤是:先"形成二、三级指标池",再作"均值、满分数量、标准差计算",进而"对有争议的指标进行讨论",并"重复以上步骤,直到专家意见趋于一致或达成共识"。研究以此设计三轮迭代:第0轮从政策文本、理论维度与首轮数据诊断汇集指标编制问卷;首轮各专家完成重要性评分、维度相关性评分与增删建议,温度参数设为0.3~0.5;第二轮各专家在群体参照下重新评分,聚焦争议指标(CV>0.25)计算W判收敛;第三轮当W<0.5或$p \geq 0.05$时启动,对高离散度指标作终裁。

(四)AI专家场景下的一致性检验方法适配

研究沿用经典检验:Kendall's W检验评分一致性(W≥0.5且$p < 0.05$)、变异系数(CV≤0.25)、内容效度指数(I-CVI≥0.78)、Cronbach's α信度(α≥0.7)。考虑AI与人类之别,另增评分偏差一致性、观点多样性指数(ODI)、跨轮评分稳定性三项专属检验。尤须强调,AI产出的指标体系须经真实人类专家验证,呼应"发挥外部利益相关者的社会监督作用,对于提升高校绩效至关重要"的共识。

四、新一轮专业质量评价标准的构建与解读

(一)新一轮评价指标体系的总体架构

新一轮评价体系的核心变革,是打破首轮"专业基础—教学运行—发展成效"的线性三阶框架,重构为"定位—教学条件—人才培养—师资—产教融合"五个评价维度,并以"动态监测"板块对规模类数据实行"监测不评价"。其重构在理念上承接CIPP"评价最重要的目的不是证明,而是改进"之旨,在机制上引入PDCA思想,经"监测预警—评估诊断—整改销号—复评确认"的闭环推动质量螺旋上升。体系设5个评价维度、23项评价指标,专业指标达标率不低于80%方为通过。依"删除—修订—保留—新增"四类处理:删除区分度

为零的失效指标,如“名师、工作室或名团队”“数字资源”;修订调整标准值者,如生师比由 35:1 收紧至 25:1、“双师型”占比由 40% 提至 50%、合作企业由“规上企业”调为“合规企业”;保留仍具区分效能者,如对口就业率、教师企业实践时长;新增填补政策缺口者,如课程体系(专业能力图谱)、内容先进、教师能力、校企共研、技术服务。

(二)新一轮评价关键改进内容的专项解读

关键改进集中于四端。其一,补强“课程与教材建设”:“人才培养”维度增设“课程体系”指标,要求构建专业能力图谱并完成合理性评价,响应教育部“绘制课程能力图谱”之求,并设“内容先进”指标纳入新标准、新工艺、新技术。其二,独立并强化“产教融合”:将校企合作独立为一级维度,设“企业教学”等六项,覆盖师资双向流动、资源联合开发、技术联合攻关诸环节。其三,师资评价由“静态资质”转向“动态能力”:新增“教师能力”落实“研究制定教师能力清单”之策,并将“企业实践”细化为每年有效实践时长不低于 30 天的可核验指标。其四,发展成效兼顾过程与结果:就业质量拆为初次就业率(≥60%)与对口就业率(≥50%)双指标,并以“监测不评价”处置学生规模,诚如朱冬玲等所言,“质量就是对标准的适切度……标准分为内适性标准和外适性标准”,评价由“一次性判断”转向“持续性治理”。

(三)指标体系的合理性验证:AI 辅助德尔菲法的共识度与信效度检验

AI 专家团队由 8 个角色的 Persona 构成,混用 DeepSeek、GPT、Gemini 等基模以降低同源趋同之险,经四节点完成研讨,三轮参与率均为 8/8。以 Kendall's W 检验评分一致性:W 值由首轮 0.4076 升至第二轮 0.5014,突破 0.5 的“良好共识”阈值,两轮 p 值均小于 0.001;第三轮为残留高分歧指标的定性终裁。三轮演化呈“分散—收敛—共识”的典型德尔菲趋势。

从单项指标离散程度看,CV≤0.25 的高度一致指标由首轮 21 项(67.7%)增至第二轮 23 项(74.2%),经第三轮终裁后全部 31 项达成共识(100%);CV>0.35 的高度分歧指标由首轮 4 项连降至第三轮 0 项。对第二轮残留的两项高分歧指标,第三轮组织 8 位专家独立终裁:“学生规模”的“监测不评价”与“科研与社会服务”改为“技术服务”,均获 8/8(100%)确认。分歧之根在于角色视角与概念理解之异,定稿以“技术服务”替代“科研”的措辞精确化解了这一概念错配。以 Cronbach's α 检验内部一致性信度:整体量表 α 由首轮 0.672 升至第二轮 0.800,达 α≥0.7 的可接受标准,表明经反馈修正后专家群体的评判趋于稳定一致;分维度子量表因题项数较少,α 低于整体量表,属正常表现。

综而论之,新一轮体系在四端获得论证:政策合规性源于对政策文本的系统对标;理论充分性源于 CIPP 决策导向模式、内适性与外适性双度评价与多元共治理论的支撑;实践可操作性源于对 23 个专业实证数据的回应;AI 赋能特色则体现为三轮收敛、整体信度达标(α:0.672→0.800)的人机协作范式。

五、结语

AI 赋能的人机协作模式,能在保全专家研讨方法学优势的前提下提升标准制定的科学性与效率;以“定位—教学条件—人才培养—师资—产教融合”五维框架重构标准、并嵌入“监测—评估—整改—复评”的 PDCA 闭环,能弥合“可评性”与“诊断性”的张力,使评价标准成为穿透文件层、进入实施层的诊断工具。由此可得三点启示:标准修订应以最新政策文本为标尺,建立“政策合规驱动”的更新机制;标准设定应参照院校实证分布,淘汰“失效指标”、改造“天花板指标”;应将数据采集内嵌于日常教学管理,以持续改进闭环替代一次性合规审查。说到底,标准之事终归关乎人——评价的终点不是一纸等级,而是教育生态中每一个具体的、生长着的生命。唯有教师、学生、专业、产业诸要素如营养素群般协同涵养,评价标准方能由冰冷的尺度化为一片孕育人性善美的绿色生态,让“教学何以成为美好事物的经历”。本研究亦坦陈局限:AI 路径样本规模有限(8 个专家角色、单校 23 个专业),推广的稳健性尚待验证;人类情感、价值判断等复杂维度 AI 尚难企及,未来需深化“何时用 AI、何时用 HI、何时需二者协同”的边界划定。

参考文献:

[1]何苗苗. 高等职业教育专业质量评价进展与路向研究[J]. 教育科学论坛,2025(3):18-24.

[2]李艳娥,吴勇. 基于适应性提升的高职专业人才培养质量标准和评价研究[J]. 广东农工商职业技术学院学报,2022,38(4):49-53.

[3]教育部,国家发展改革委,工业和信息化部,等. “人工智能+教育”行动计划(教科信〔2026〕1 号)[Z]. 2026.

[4]四川省教育厅. 四川省教育数字化行动计划(2025—2027 年)[Z]. 2025.

[5]教育部职业院校信息化教学指导委员会. 职业院校人工智能应用指引(2026 版)[Z]. 2026.

[6]罗勇,周玉梅,程晨林,等. 高职院校内部专业评价指标体系及其评价标准研究[J]. 宁波职业技术学院学报,2022,26(5):57-62.

[7]满冬,张波,李华勇. 高职院校专业质量诊断与评价指标体系及实施路径研究[J]. 现代职业教育,2023(36):38-41.

[8]四川省教育厅. 四川省高等职业学校办学能力评价实施方案(2025—2030 年)[Z]. 成都:四川省教育厅,2025.

[9]教育部. 高等学校人才培养质量保障体系建设指南[Z]. 2025.

[10]教育部. 关于深化职业教育教学关键要素改革的意见(教职成〔2026〕1 号)[Z]. 2026.

[11]徐春艳. 多元共治视域下高职院校专业建设外部质量监督与评价协同机制研究[J]. 常州信息职业技术学院学报,2025,24(1):8-12.

[12]刘静. 高职院校高水平专业群评价指标体系构建[J]. 创新创业理论研究与实践,2025(10):4-6.

[13]朱冬玲,周秀英,唐世文,等. 培养目标达成度和职业岗位匹配度视域下的专业人才培养质量评价[J]. 广州城市职业学院学报,2022,16(2):88-93.

[14]刘铁芳,黄鑫. 教学何以成为美好事物的经历[J]. 中国教育学刊,2014(7):33-38.

Reconstruction of Higher Vocational Program Quality Evaluation Standards Driven by Agent-based Delphi Method

YANG Zhuo-dong
(Bazhong Vocational and Technical College, Bazhong Sichuan 636000, China)

Abstract: To address two pathologies exposed in the first-round trial of program quality evaluation at a higher vocational college, missing dimensions in policy responsiveness and failed discrimination in practical effectiveness, this study introduces artificial intelligence to empower the revision of evaluation standards. Following a "policy benchmarking — problem diagnosis — AI empowerment — standard reconstruction" pathway, it conducts a dual review of the existing standards against both the latest policy texts and implementation data from 23 programs, and designs a two-tier architecture of "8 AI expert roles + human review and adjudication". Through three iterative rounds of an AI-assisted Delphi method, the revision is completed and its consensus and reliability are verified. The original linear three-stage framework is thereby reconstructed into a five-dimensional system of "positioning, teaching conditions, talent cultivation, faculty, and industry-education integration", in which scale-related data are "monitored but not evaluated", and a "monitoring — assessment — rectification — re-evaluation" PDCA closed loop is embedded. The study proposes a "human-AI collaboration" paradigm for standard-setting and a "data-driven PDCA (D-PDCA)" paradigm for governance, offering a methodological reference for comparable higher vocational colleges.

Key words: artificial intelligence; program quality evaluation; evaluation standards; Delphi method; PDCA closed loop

(责任编辑:桂杉杉)

(上接第 84 页)

[4]安徽省教育厅. 关于 2023 年安徽省高等职业教育"双师型"教师认定结果的公示[EB/OL]. (2023-12-31). http://jyt.ah.gov.cn/public/7071/40673558.html.

[5]甘肃省教育厅. 甘肃省职业教育"双师型"教师认定实施办法[EB/OL]. (2023-1-15)[2023-1-30]. https://zhijiao.eol.cn/2023/01/30/9925750.html.

[6]福建省教育厅. 福建省职业教育"双师型"教师认定实施办法(试行)[EB/OL]. (2023-12-7). https://jyt.fujian.gov.cn/xxgk/zfxxgkzl/zfxxgkml/gfxwj/202312/t20231212_6329924.htm.

The Current Status, Challenges, and Recommendations for the Implementation of "Double-qualified" Teacher Certification in Vocational Colleges in China

KANG Fei, JIA Qian-qian, LIU Jing-xue
(Guidaojiaotong Polytechnic Institute, Shenyang Liaoning 110023, China)

Abstract: After the introduction of the national "double-qualified" teacher recognition standards, various provinces (autonomous regions and municipalities) have successively carried out the recognition work for "double-qualified" teachers. Although the recognition work has achieved remarkable results, there are still issues such as the need to strengthen the management system and mechanism for recognition, improve the adaptability of provincial recognition standards, enhance the construction of information platforms, strengthen incentive mechanisms, and increase the intensity of training and practice for "double-qualified" teachers. This paper summarizes and analyzes the relevant issues and proposes policy recommendations from aspects such as "improving institutional mechanisms", "guiding classification recognition", "enhancing the informatization level of recognition", "improving incentive mechanisms", and "strengthening training and cultivation efforts", to help promote the recognition work for "double-qualified" teachers in vocational education.

Key words: "double-qualified" teachers; implementation status; policy recommendations

(责任编辑:陈思婷)