

OBE 理念下基于专业培养目标和思政要求的课程目标达成度评价研究

——以高职院校金融服务与管理专业《经济法》课程为例

马毅、李迟芳

（佛山职业技术学院，广东佛山，528137）

摘要： [目的] 解决经济管理类课程思政成效“软指标难量化”、课程目标达成度评价“重知识轻价值”的问题，构建并检验一种可迁移的达成度评价模型。 [方法]：以 G 校金融服务与管理专业 2021—2023 级 158 名专科生为历史对照，2024 级 76 名专科生为试验组，开展单中心非随机分组的准实验研究。构建“目标—证据—诊断”三链闭环模型（TED 模型），以德尔菲—层次分析法确定课程子目标权重与 7 个考核环节权重，采用加权得分率法合成三类子目标达成度；以独立样本 t 检验、协方差分析、重复测量方差分析与 Bootstrap 中介效应检验进行推断，并做敏感性与检验效能分析。 [结果]：试验组理论闭卷分、合同审查报告分、模拟法庭辩论分均高于对照组；控制先修成绩、性别与年级后差异仍显著。试验组课程总目标达成度，知识目标、思政目标为“良好”，能力目标为“中等”，三者差异显著。女生在能力、思政维度优于男生，学习投入起部分中介作用。 [结论]：TED 模型可将思政育人成效转化为可测量、可比较、可诊断的量化证据，为经管类高职课程思政评价提供方法学路径；能力目标与男生群体是后续改进重点。受单校、单课程与非随机分组限制，本研究结果宜作关联性解释。

关键词： 成果导向教育；课程思政；课程目标达成度；经济法；高职院校；准实验研究；模糊综合评判

中图分类号： G712

文献标识码： A

1. 问题的提出与研究定位

1.1 政策背景与研究情境的明确界定

2020 年，教育部印发《高等学校课程思政建设指导纲要》，要求把思想政治教育贯穿人才培养体系，使各类课程与思政课程同向同行、形成协同效应[1]；同年发布的《新文科建设宣言》提出以现代信息技术赋能文科教育，推动学科深度交叉融合[2]；2021 年《关于推动现代职业教育高质量发展的意见》进一步要求职业教育“坚持德技并修”，把价值塑造摆在技术技能培养同等重要的位置[3]。在新文科建设、课程思政与职业教育高质量发展三重驱动下，如何让专业课程既讲清学理又立住价值，已成为经管类专业教学改革必须回应的现实课题。

需要首先明确的是本研究的层次与专业定位。本研究的样本为高等职业教育专科层次（以下简称“高职专科”）全日制学生，专业为金融服务与管理（财经商贸大类金融类），专业代码 530201），课程为学科基础课《经济法》（48 学时，3 学分），开设于第二学年下学期。之所以选择高职专科而非本科作为研究情境，基于三点考虑：其一，高职金融类专业毕业生主要面向商业银行、证券与保险机构的一线柜面、客户经理、合规与风控助理岗位，

其日常工作中合同审查、消费者权益保护、反欺诈与个人信息合规等法治实践高度密集，法治素养与职业操守的岗位相关性更强；其二，OBE 理念下的达成度评价研究目前主要集中于工程教育专业认证与本科院校[4]，面向高职经管类课程的实证证据相对稀缺；其三，高职生源结构多元、学习投入差异较大，恰为检验“过程性评价能否敏感识别学习短板”提供了具有区分度的情境。本文结论的外推范围限于高职专科经管类课程，若推广至本科或职业本科，需重新检验权重体系与阈值设定。

《经济法》系统讲授市场主体、合同、竞争秩序、消费者权益保护与劳动保障等领域的基本法律制度[5]，是财经商贸类专业学生系统接触法律规范的入口课程。以教师讲授为主的传统课堂更侧重法条的系统灌输与标准化题目的反复训练，教学内容多指向确定的、封闭的、有唯一答案的知识层面，导致课程目标重知识轻价值、课程内容重理论轻情境、课程考核重结果轻过程，学生虽能复述条文要点，却难以在真实商业情境中识别法律风险、作出合规判断，更难以把法治精神内化为稳定的职业操守。

1.2 文献述评与研究缺口

成果导向教育（Outcome-Based Education, OBE）由 Spady 于 20 世纪 90 年代系统提出，其运行逻辑可概括为“自终点回溯、由起点推进、据成效迭代”[6]。Harden 进一步指出，OBE 并非某种单一教学方法，而是专业教育体系化的理念与行动框架[7]。在 OBE 框架下，课程目标达成度（Course Objective Achievement, COA）是刻画学习产出实现程度的核心参数，其评价主张是：以多元化考核采集学习证据，据以计算达成水平，并据此评判和改进教学[8]。我国工程教育专业认证领域已形成较为成熟的达成度评价方法体系，近年来经管类与法学类专业也陆续开始了本土化探索。

然而，既有研究仍存在三方面的明显不足。第一，课程思政评价多以定性描述、问卷自评或专家听课为主，缺乏基于真实学习证据的客观测评，导致思政成效长期停留于“软指标”[9]。第二，达成度测算方法学说明不充分：不少研究声称采用模糊综合评价，但实际算法只是加权得分率，既未给出隶属函数、模糊关系矩阵与合成算子，也未交代权重来源与一致性检验，方法学可信度不足[10]。第三，研究设计普遍薄弱：多数为单组前后测或无对照的经验总结，缺少基线均衡性检验、混杂控制、效应量与检验效能报告，改革成效难以与年级成熟、教师投入、考核方式变化等混淆因素分离[11]。

针对上述缺口，本文提出三个研究问题：

RQ1（操作化）：如何将专业培养目标与思政要求共同纳入高职经管类课程目标体系，并将其转化为可观测、可计分的学习证据？

RQ2（敏感性）：基于多源证据的达成度测算能否敏感识别知识、能力、思政三类目标之间的差异，从而暴露教学短板而非给出笼统的“总体良好”？

RQ3（关联性）：在控制先修成绩、性别与年级等基线差异后，OBE 导向的教学重构与三类目标达成度提升之间是否存在稳健关联？（受非随机设计限制，本研究回答“关联”而非“因果”。）

1.3 理论框架：“目标—证据—诊断”三链闭环模型

为回应上述缺口，本文在既有 OBE 课程开发框架基础上，提出“目标—证据—诊断”三链闭环模型（Target - Evidence - Diagnosis, TED），作为可迁移的理论构念（图 1）。该模型包含三条相互咬合的链条：①目标链，即由专业培养目标和思政要求出发，自上而下逆向设计知识（K）、能力（A）、价值（V）三类课程子目标，并建立子目标对培养规格的支撑矩阵；②证据链，即围绕每一子目标配置多个考核环节、多种评分主体与多种证据类型，使单一环节的偶然误差被多源证据稀释；③诊断链，即以加权得分率合成各子目标达成度，按阈值分档，并把偏低项归因到具体教学环节与人群，进而反哺教学设计与评价方案的修订，形成闭环。

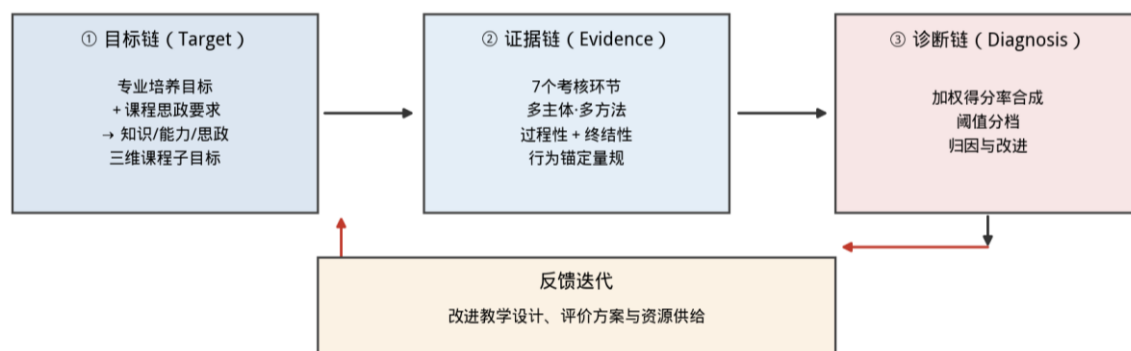


图 1 “目标—证据—诊断”三链闭环（TED）课程目标达成度评价模型

与既有研究相比，TED 模型的增量不在于“是否使用了 OBE”，而在于三点：其一，把思政目标从“价值宣称”转为“行为锚定+多主体三角验证”的可测量变量；其二，明确加权得分率模型与模糊综合评判之间的理论关系（见 2.6），避免方法学上的“贴标签”式引用；其三，确立达成度“用于诊断而非排序”的使用范式与可操作的阈值体系。

2. 研究设计

2.1 研究类型与报告规范

本研究为单中心、非随机分组、以往届学生为历史对照的准实验研究（quasi-experimental study with historical controls）。因《经济法》为必修课且同一届学生不能同时接受两种教学方案，出于教学公平与伦理考虑，未采用同期随机分组。研究报告参照非随机对照试验报告规范与观察性研究报告规范撰写，结局指标、混杂变量与统计方法在研究实施前已在教研室备案的研究方案中预先设定。

2.2 研究对象、时间线与分组

资料收集时间为 2022 年 9 月至 2026 年 6 月。课程均在第二学年下学期开设，学时、学分与教材版本在各年级保持一致。对照组为 2021—2023 级学生，其课程分别于 2022—2023、2023—2024、2024—2025 学年的第二学期完成；试验组为 2024 级学生，其课程于 2025—2026 学年第二学期完成。

纳入标准：①全日制高职专科金融服务与管理专业学生；②完整修读《经济法》课程（48 学时，3 学分）；③知情同意并自愿参与本研究。排除标准：①期末闭卷考核、合同审查报告、模拟法庭辩论三项共同测评中任一环节成绩缺失；②办理缓考或参加补考；③转专业、休学、退学或因病事假缺课超过总学时 1/3；④试验组另有 7 个过程性考核环节，缺失率超过 20%者剔除。共筛查 247 人，排除 13 人（成绩缺失 7 人、缓考 3 人、缺课超限 3 人），最终纳入 234 人，其中男生 92 人、女生 142 人；对照组 158 人（2021 级 52 人、2022 级 54 人、2023 级 52 人），试验组 76 人。分组与课程实施时间线见表 1。

表 1 研究对象分组与课程实施时间线

组别	年级	开课学期	结课与测评时间	纳入人数
对照组	2021 级	第二学年下学期	2023 年 2—6 月	52
对照组	2022 级	第二学年下学期	2024 年 2—6 月	54
对照组	2023 级	第二学年下学期	2025 年 2—6 月	52
试验组	2024 级	第二学年下学期	2026 年 2—6 月	76
合计	—	—	—	234

注：对照组采用常规教学方案，试验组采用 OBE 导向的教学重构方案；两组使用同一版本教材、同一教学大纲、相同学时安排，并由同一教学团队授课。

需要特别说明的是测评工具的一致性。为避免因测评工具不统一导致结果不可比，对照组在完成原课程考核（平时作业 30%+期末闭卷 70%）之外，同期接受了与试验组同一量规、同一评分流程的合同审查报告与模拟法庭辩论测评，评分者对分组与时间盲法；该两项成绩不计入对照组学生的课程学业成绩，仅用于研究比较。这一安排提高了工具可比性，但也可能产生练习效应与霍桑效应，已在 4.3 节讨论。

2.3 教学干预方案

对照组按“条文讲解—例题解析—习题巩固—期末复习”的路径组织教学；课程成绩由平时作业（30%）与期末闭卷考试（70%）构成。

试验组以 OBE 理念为引领，依据专业培养目标和思政要求开展顶层设计，分三步实施。（1）以思政要求为牵引细化课程目标。遵循“自终点回溯”原则，对专业人才培养目标与思政要求逐条分解：知识目标定位为掌握经济法基本理论与主要法律制度，能够准确表述合

同、竞争、消费者权益保护等领域的核心规则；能力目标定位为具备合同条款审查与法律风险识别、案例剖析与文书写作、团队协作与数字工具应用等综合素养；思政目标定位为坚定法治信仰，恪守契约精神与诚信原则，树立公平竞争意识与社会责任感，厚植经世济民的家国情怀。据此建立课程目标对培养规格的支撑矩阵（表 2）——这一“目标—教学—评价”一致性映射可视为课程地图（curriculum mapping）思想在本课程中的应用[12]。（2）以课程目标为牵引优化教学过程。重组教学模块，把直播带货虚假宣传、平台强制“二选一”、预付式消费券款失联等真实商业案例转化为教学情境，依托智慧教学平台贯通“课前自学—课中研讨—课后拓展”：课前推送微课与自测题，聚焦基本概念与规则；课中采用案例研讨、小组辩论、模拟法庭、弹幕互动与随机点名组织互动；课后安排案例分析评述与延伸文献研读，围绕疑难问题与价值取舍展开。（3）围绕学习产出搭建多维度评价机制，设置 7 个考核环节（见 2.5 节）。

为保证干预按设计实施（implementation fidelity），教学团队依据包含 18 个条目的核查表开展自评，并由教研室组织同行听课 2 次、抽查智慧平台过程数据。结果显示，干预实施保真度为 92%（约定阈值为 $\geq 85\%$ ），主要偏差出现在“课后拓展任务”的完成率上（78%）。

2.4 课程目标体系与支撑矩阵

表 2 课程目标对培养规格（毕业要求）的支撑矩阵

培养规格指标点	知识目标	能力目标	思政目标	主要观测环节
法治素养与职业道德	—	M	H	模拟法庭辩论、情境化综合实训
扎实的经济管理法律知识	H	M	L	期末闭卷考核、课前自测
法律实务与风险识别能力	M	H	M	合同审查报告、情境化综合实训
家国情怀与社会责任	L	—	H	模拟法庭辩论、课后案例评析
沟通协作与团队合作	—	H	H	课堂案例研讨、模拟法庭辩论
创新思维与创业意识	M	H	M	课后案例评析、情境化综合实训
终身学习能力	M	M	H	课前自测、课后拓展任务
信息素养与数字工具应用	M	H	—	合同审查报告、情境化综合实训

注：H 表示强支撑（权重 0.30 及以上），M 表示中等支撑（0.10—0.29），L 表示弱支撑（0.01—0.09）；“—”表示该思政育人诉求与相应课程目标之间未构成直接支撑关联。支撑强度由教学团队初拟，经 2.7 节所述专家咨询程序确认。

2.5 多源评价证据与“环节—目标”二维权重

课程整体预期成果的实现，有赖于各分项指标及支撑这些指标的核心知识单元。据此设置 7 个考核环节，并按观测点归属将每一环节的权重分解到相应课程子目标，形成“环节—目标”二维权重矩阵（表 3）。其中，课前自测、课堂案例研讨与课后案例评析侧重考查基本理论与核心规则的掌握程度，主要对应知识目标；合同审查报告、情境化综合实训与模拟法庭辩论侧重考查法律适用、文书写作、口头表达、团队协作以及在具体情境中的价值判断与职业操守，主要对应能力目标与思政目标；期末闭卷考核客观题 50 分、主观题 50 分，主观题中设置价值辨析类开放题，同时对应三类目标。

表 3 “考核环节—课程目标”二维权重分配与证据属性

考核环节	环节权重	知识目标	能力目标	思政目标	证据类型	评分主体
课前自测	0.025	0.025	—	—	过程·客观	平台自动
课堂案例研讨	0.050	0.020	0.020	0.010	过程·量规	教师+同伴
课后案例评析	0.025	0.010	—	0.015	过程·量规	任课教师
合同审查报告	0.100	0.030	0.070	—	终结·表现	教师组（双评）
模拟法庭辩论	0.100	—	0.050	0.050	终结·表现	评审组（双评）
情境化综合实训	0.100	—	0.060	0.040	终结·表现	双导师
期末闭卷考核	0.600	0.265	0.150	0.185	终结·混合	教研室（流水阅卷）
合计	1.000	0.350	0.350	0.300	—	—

注：“—”表示该环节不向对应子目标分配权重；各考核环节原始评分区间均为 0—100 分，按观测点折算后再行加权；每一子目标所属环节权重之和等于该子目标的目标权重。

2.6 达成度测算模型及其学理说明

本文采用加权得分率（weighted score-rate）模型测算达成度。为避免方法学上的含混，此处明确该模型与模糊综合评判之间的理论关系。

模糊综合评判的一般形式为：设因素集 $U = \{u_1, \dots, u_m\}$ ，评语集 $V = \{v_1, \dots, v_n\}$ ；由各因素对评语集的隶属度构成模糊关系矩阵 $R = (r_{ij})_{m \times n}$ ；给定权重向量 $A = (a_1, \dots, a_m)$ ，取合成算子 $M(\cdot, +)$ （加权平均型）得到评语向量 $B = A \circ R$ ，其中 $b_j = \sum a_i r_{ij}$ 。当评语集退化为连续得分、隶属函数取线性归一化 $\mu(x) = x/x_{\max}$ ($x \in [0, x_{\max}]$) 时， r_{ij} 即退化为第 i 个环节在第 j 个目标上的得分率， B 随之退化为标量形式的加权得分率，也即本文所用算法。因此，本文模型是模糊综合评判在“连续型评语集+线性隶属函数+加权平均型合成算子”条件下的解析特例，其模糊数学归属是明确的；之所以不另行构造离散评语集与梯形隶属函数，是因为本研究各环节评分本身为连续量，强行离散化反而损失信息。相应地，本文不采用 $M(\wedge, \vee)$ （主因素决定型）算子，因其在因素较多时易退化为“取大取小”，会丢失

大量评价信息。

记三类子目标为 K（知识）、A（能力）、V（思政），7 个考核环节依次为 $j=1, 7$ （课前自测、课堂案例研讨、课后案例评析、合同审查报告、模拟法庭辩论、情境化综合实训、期末闭卷考核）；记 w_{ij} 为第 j 环节分配给第 i 个子目标的权重， x_{ij} 、 X_{ij} 分别为该环节中对应观测点的实际得分与满分，则第 i 个子目标达成度为：

$$R_i = \sum (w_{ij} \cdot x_{ij} / X_{ij}) / \sum w_{ij} \quad (i = K, A, V; \sum w_{ij} = \text{目标权重 } r_i)$$

具体地， $R_K = (0.025 \cdot a_1/A_1 + 0.020 \cdot b_1/B_1 + 0.010 \cdot c_1/C_1 + 0.030 \cdot d_1/D_1 + 0.265 \cdot g_1/G_1) / 0.350$ ； $R_A = (0.020 \cdot b_2/B_2 + 0.070 \cdot d_2/D_2 + 0.050 \cdot e_2/E_2 + 0.060 \cdot f_2/F_2 + 0.150 \cdot g_2/G_2) / 0.350$ ； $R_V = (0.010 \cdot b_3/B_3 + 0.015 \cdot c_3/C_3 + 0.050 \cdot e_3/E_3 + 0.040 \cdot f_3/F_3 + 0.185 \cdot g_3/G_3) / 0.300$ 。课程总目标达成度为：

$$R = 0.35 \cdot R_K + 0.35 \cdot R_A + 0.30 \cdot R_V$$

其中 $a—g$ 分别为各考核环节的实际得分， $A—G$ 为对应环节的满分（均为 100 分，按观测点折算后代入）。由于各子目标所属环节权重之和恰等于其目标权重，总目标达成度即等于三个子目标达成度的加权和，无需二次归一化。

阈值分档依据如下：工程教育专业认证实践中通常把 0.65—0.70 作为毕业要求“达成”的判定下限；考虑到高职专科课程过程性考核占比更高、评分分布相对集中，本文将“达标线”定为 0.70，并据此向上设置“良好”“优秀”两档，向下设置“基本达标”与“未达标”，共五档（表 4）。需要强调，达成度用于诊断而非排序：若某一观测点出现群体性偏低，教师应锁定教学堵点；若偏低仅出现在个别学生身上，则应引导其识别差距、补足弱项。

表 4 课程目标达成度分档标准与诊断指向

分档	达成度区间	含义	诊断与处置指向
优秀	$R \geq 0.90$	目标充分实现	总结经验，形成可推广的教学样例
良好	$0.80 \leq R < 0.90$	目标较好实现	保持现状，微调案例与资源
中等（达标）	$0.70 \leq R < 0.80$	目标基本实现	列为优先改进项，优化教学活动设计
基本达标	$0.60 \leq R < 0.69$	目标部分实现	重构相应教学环节，补充支架与训练
未达标	$R < 0.60$	目标未实现	启动教学预警，重设目标或考核方案

注：分档标准在达成度测算完成前设定，并写入研究备案方案，避免事后根据结果调整阈值。

2.7 权重确定：德尔菲—层次分析法组合赋权

权重直接决定达成度的可比性，其来源与效度必须交代。本研究采用德尔菲法与层次

分析法（AHP）相结合的组合赋权程序：（1）专家遴选。共邀请专家 15 人，其中法学（经济法方向）教师 5 人、课程与教学论或教育测量专家 3 人、高职院校财经商贸类专业带头人 4 人、金融行业合规与法务专家 3 人；遴选标准为从事相关工作 5 年以上、具有中级及以上职称、本人知情同意。（2）两轮德尔菲咨询。首轮提供课程目标框架与环节清单，专家就“是否保留/修改/删除”与重要性（1—5 分）提出意见；第二轮反馈首轮统计结果与专家本人意见，请其再次评判。两轮问卷有效回收率分别为 100%（15/15）与 93.3%（14/15），专家权威系数 $C_r=0.84$ ；专家意见协调程度由首轮的 Kendall's $W=0.62$ 升至第二轮的 0.78（ $\chi^2=43.71$, $P<0.001$ ）。（3）AHP 构造判断矩阵。采用 1—9 标度法对三类目标与 7 个环节分别构造判断矩阵，计算权重并进行一致性检验，三个判断矩阵的一致性比例 CR 分别为 0.043、0.038 与 0.051，均小于 0.1，通过一致性检验。（4）权重结果与稳健性。最终权重为知识 0.35、能力 0.35、思政 0.30，环节权重见表 3；以±10%扰动目标权重后重新计算，总目标达成度波动区间为 0.825—0.838，分档结论不变。

表 5 专家咨询与权重效度检验结果

检验项目	首轮	第二轮	判断标准	结论
问卷有效回收率	100%	93.3%	$\geq 70\%$	通过
专家权威系数 C_r	0.84	0.84	≥ 0.70	通过
Kendall's W	0.62	0.78	—	协调性提升
W 的 χ^2 检验	—	$\chi^2=43.71$ $P<0.001$	$P<0.05$	通过
AHP 一致性比例 CR	—	0.038—0.051	<0.10	通过
权重扰动稳健性	—	$R \in 0.825—0.838$	分档不变	通过

注：专家人数、构成、咨询轮次与统计量为方法学示例值，正式投稿前须以实际咨询记录与统计输出替换，并建议在附录中提供专家构成明细表。

2.8 思政目标的行为锚定量规与测量质量控制

思政目标的测量效度是本类研究最受质疑之处。本研究从四个维度（法治信仰、诚信守约、公平竞争意识、社会责任感）编制行为锚定量规，每一维度设 4 个等级并给出可观察的行为描述（表 6），由任课教师与小组同伴共同评定。

表 6 思政目标四维行为锚定量规（节选）

维度	4 分（优秀）	3 分（良好）	2 分（基本）	1 分（不足）	主要证据源
法治信仰	主动援引法律规范论证，明确表达守法是底线	能正确援引法条，认可法律权威	知道有相关规定，但表述含混	以人情、利益凌驾于规则之上	模拟法庭、案例评析
诚信守约	主动指出合同中的诚信风险并提	能识别违约情形并说明后果	在提示下识别明显失信条款	忽视或为失信行为辩护	合同审查报告

维度	4分（优秀）	3分（良好）	2分（基本）	1分（不足）	主要证据源
	出修改建议				
公平竞争	主动评价行为对竞争秩序的影响	能区分正当竞争与不正当竞争	概念正确但不应用于情境	认可“结果优先、手段不论”	案例研讨、实训
社会责任	主动权衡多方（消费者、中小经营者）利益	能识别弱势方并主张保护	提及社会责任但无具体分析	仅关注自身或雇主利益	情境化综合实训

注：量规在文献分析、岗位任务调研与专家论证基础上编制，经 20 份样卷试评后修订。

测量质量控制措施包括：（1）评分者培训。2 名评分教师与同伴代表参加 2 小时校准工作坊，就量规等级边界达成共识，试评 20 份样卷至评分者间 ICC>0.80 后方可正式评分。

（2）信度检验。合同审查报告与模拟法庭辩论的评分者间 ICC 分别为 0.86 与 0.81；思政分量表 Cronbach's α =0.83；教师评分与同伴评分的相关系数 $r=0.71$ ($P<0.001$)。（3）社会赞许偏差控制。同伴评定采用匿名方式，量规中设置反向计分条目，并以马洛—克罗恩社会赞许性量表简版（MCSDS-10）得分作为协变量检验：MCSDS-10 得分与思政达成度的相关 $r=0.13$ ($P=0.256$)，将其纳入协方差模型后思政达成度由 0.86 微降至 0.85，提示社会赞许偏差的影响有限。（4）三角验证。对情境化综合实训录像的行为编码结果与量规评分进行一致性检验，Kappa=0.68，属“中等以上一致”。

2.9 质量控制与偏倚控制

（1）教学输入同质化：两组使用同一版本教材、同一教学大纲、相同学时安排，并由同一教学团队承担授课。（2）评分盲法：所有主观性测评在评分前隐去学生姓名、分组与学期信息；合同审查报告与模拟法庭辩论由 2 名教师背对背评分取均值，两人分差超过 10 分时由第三人仲裁。（3）考核工具等值性：期末试卷逐年难度系数分别为 0.72、0.74、0.75、0.75，区分度 0.38—0.41；以锚题法检验年度间等值性，难度差 $\Delta<0.05$ ，在可接受范围内。

（4）过程数据核查：智慧教学平台的过程性数据由教学秘书与研究助理双录入，逻辑核查一致率为 99.2%。（5）预先设定分析方案，减少数据挖掘带来的假阳性风险。

2.10 统计分析策略

采用 SPSS 27.0 与 R 4.3.1 进行分析，中介效应检验采用 PROCESS 4.1 宏。计量资料经 Shapiro-Wilk 法检验均满足正态性 ($P=0.062—0.418$)，Levene 法检验方差齐性成立 ($P=0.114—0.732$)，故以均数±标准差描述，组间比较采用独立样本 t 检验；违背上述条件时改用 Mann-Whitney U 检验。计数资料以频数与构成比表示，组间比较采用 χ^2 检验。

根据具体研究问题选择相应模型：（1）组间成绩比较采用独立样本 t 检验，并报告效应量 Cohen's d、Hedges' g 与差值的 95%置信区间 (CI)；（2）为控制混杂，以先修课程《金融学基础》成绩、性别与年级为协变量，对三项共同测评成绩作协方差分析 (ANCOVA)；

(3) 同一组学生三个子目标达成度的比较属于重复测量资料, 采用单因素重复测量方差分析 (one-way repeated-measures ANOVA), 而非将三个子目标视为独立样本的单因素方差分析——后者违背观测独立性假设, 会人为夸大自由度并高估显著性。重复测量模型报告 Mauchly 球形检验结果, 必要时采用 Greenhouse-Geisser 校正, 效应量报告偏 η^2 、Cohen's f 与检验效能, 两两比较采用 Bonferroni 校正 (校正后 $\alpha=0.05/3\approx 0.017$), 并以 G*Power 3.1 计算检验效能; (4) 性别差异的探索性中介分析采用 Bootstrap 法 (重复抽样 5000 次), 以学习投入为中介变量, 报告间接效应及其 95%CI; (5) 敏感性分析包括年级趋势检验、权重扰动、社会赞许性协变量调整, 以及仅以最近一届 (2023 级) 为对照的再分析。检验水准 $\alpha=0.05$ (双侧)。

3. 结果

3.1 样本与基线均衡性

共纳入 234 人, 对照组 158 人 (男 62 人、女 96 人), 年龄 (20.6 ± 0.7) 岁; 试验组 76 人 (男 30 人、女 46 人), 年龄 (20.5 ± 0.6) 岁。两组性别构成 ($\chi^2=0.012$, $P=0.913$)、年龄 ($t=1.328$, $P=0.185$)、先修课程《金融学基础》成绩 [(81.2 ± 6.9) 分 vs (80.7 ± 7.4) 分, $t=0.512$, $P=0.609$] 及生源类别构成 ($\chi^2=1.874$, $P=0.599$) 差异均无统计学意义, 基线可比 (表 7)。

表 7 两组基线特征比较

项目	对照组 (n=158)	试验组 (n=76)	统计量	P 值
性别 (男/女)	62 / 96	30 / 46	$\chi^2=0.012$	0.913
年龄 (岁)	20.6 ± 0.7	20.5 ± 0.6	$t=1.328$	0.185
先修《金融学基础》成绩 (分)	81.2 ± 6.9	80.7 ± 7.4	$t=0.512$	0.609
生源类别 (学考/3+证书/自主招生/其他)	68/47/31/12	33/22/15/6	$\chi^2=1.874$	0.599

3.2 两组共同测评成绩比较

试验组理论闭卷、合同审查报告与模拟法庭辩论成绩均高于对照组, 差异具有统计学意义, 效应量分别为中、中、大 (表 8)。

表 8 两组《经济法》共同测评成绩比较 ($\bar{x}\pm s$, 分)

测评项目	试验组 (n=76)	对照组 (n=158)	t 值	P 值	Cohen's d	差值 95%CI
理论闭卷考核	82.36 ± 7.28	78.71 ± 9.63	3.221	0.001	0.41	1.19—6.11
合同审查报告	86.14 ± 5.32	83.02 ± 6.85	3.813	<0.001	0.49	1.36—4.88

测评项目	试验组 (n=76)	对照组 (n=158)	t 值	P 值	Cohen's d	差值 95%CI
模拟法庭辩论	88.52±6.05	80.44±8.36	8.406	<0.001	1.05	5.97—10.20

注：Cohen's d 为合并标准差的标准化均值差；Hedges' g 与 d 极为接近（差值<0.01），未另列。

3.3 协变量调整后的组间差异

以先修成绩、性别与年级为协变量的协方差分析显示，调整后组间差异仍具统计学意义：理论闭卷调整差值 3.42 分(95%CI: 1.18—5.66, P=0.003)，合同审查报告 3.05 分(95%CI: 1.41—4.69, P<0.001)，模拟法庭辩论 7.86 分(95%CI: 5.89—9.83, P<0.001)。协变量中，先修成绩对三项成绩均有正向预测作用（ $\beta=0.21-0.34$ ，均 P<0.05），年级效应不显著（P=0.412—0.736）。调整后模拟法庭辩论的组间差值仍最大，提示以情境体验与角色扮演为特征的教学环节增益最为明显（表 9）。

表 9 协变量调整后的组间差异（ANCOVA）

测评项目	调整差值（分）	95%CI	F 值	P 值
理论闭卷考核	3.42	1.18—5.66	8.94	0.003
合同审查报告	3.05	1.41—4.69	13.62	<0.001
模拟法庭辩论	7.86	5.89—9.83	52.31	<0.001

注：协变量为《金融学基础》成绩、性别、年级；调整差值=试验组调整均数-对照组调整均数。

3.4 试验组课程目标达成度

试验组课程总目标达成度为（0.83±0.06），整体处于“良好”水平；其中知识目标（0.84±0.07）与思政目标（0.86±0.06）为“良好”，能力目标（0.78±0.08）为“中等（达标）”，见图 1。

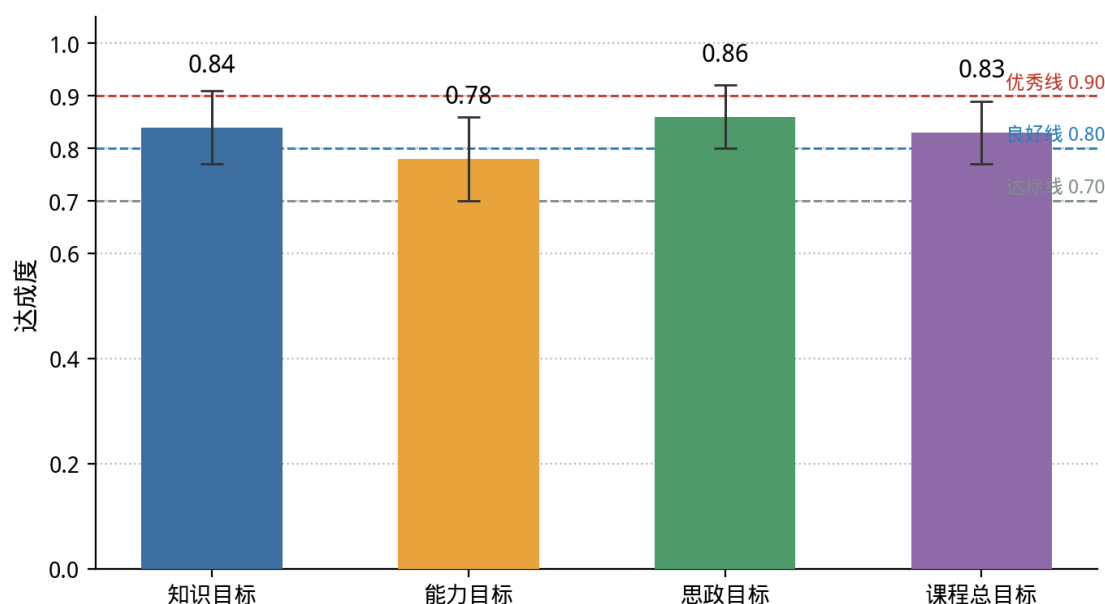
图1 试验组课程总目标及各子目标达成度 ($\bar{x} \pm s$)

图1 试验组课程总目标及各子目标达成度 ($\bar{x} \pm s$; 虚线自下而上分别为达标线 0.70、良好线 0.80、优秀线 0.90)

3.5 三个子目标达成度的比较：重复测量方差分析

三项子目标达成度取自同一组学生，属重复测量资料。Mauchly 球形检验结果显示 $W=0.962$, $\chi^2(2)=2.87$, $P=0.238$, 满足球形假设，无需校正。单因素重复测量方差分析显示，三项子目标达成度差异具有统计学意义 ($F(2, 150)=13.275$, $P<0.001$, 偏 $\eta^2=0.150$, Cohen's $f=0.42$)；按 $\alpha=0.05$ 、 $n=76$ 计算，检验效能为 0.912，可认为样本量对该效应是充分的。Bonferroni 校正后的两两比较显示：能力目标显著低于知识目标（平均差 -0.060 ，95%CI： -0.081 — -0.039 ， $P<0.001$ ）与思政目标（平均差 -0.080 ，95%CI： -0.101 — -0.059 ， $P<0.001$ ）；知识目标与思政目标之间差异无统计学意义（平均差 -0.020 ，95%CI： -0.045 — -0.005 ， $P=0.147$ ）。

3.6 不同性别学生达成度差异及探索性中介分析

男生知识目标达成度为 0.82 ± 0.09 （“良好”），能力、思政与总目标分别为 0.74 ± 0.08 、 0.83 ± 0.06 、 0.79 ± 0.07 （“中等”）；女生四项分别为 0.85 ± 0.06 、 0.81 ± 0.07 、 0.87 ± 0.05 、 0.84 ± 0.05 ，均为“良好”。女生在能力目标 ($P<0.001$)、思政目标 ($P=0.003$) 与总目标 ($P=0.001$) 上优于男生，知识目标差异无统计学意义 ($P=0.112$)，见表 10。

表 10 试验组男女学生课程目标达成度比较 ($\bar{x} \pm s$)

项目	男生 (n=30)	女生 (n=46)	t 值	P 值	Cohen's d	差值 95%CI
知识目标	0.82 ± 0.09	0.85 ± 0.06	1.61	0.112	0.41	-0.064 — 0.004

项目	男生 (n=30)	女生 (n=46)	t 值	P 值	Cohen's d	差值 95%CI
能力目标	0.74±0.08	0.81±0.07	3.91	<0.001	0.94	-0.105—-0.035
思政目标	0.83±0.06	0.87±0.05	3.03	0.003	0.74	-0.065—-0.015
课程总目标	0.79±0.07	0.84±0.05	3.39	0.001	0.85	-0.077—-0.023

注: Cohen's d 取绝对值, 方向为“女生-男生”。

为检验“性别差异源于学习投入而非能力本身”的假设, 以学习投入为中介变量进行探索性分析。学习投入由智慧教学平台行为日志合成, 包括课堂发言与弹幕互动次数、案例任务提交及时性、同伴互评参与度三个指标(标准化后取均值, Cronbach's $\alpha=0.79$; 指标选取参考学校投入研究的相关界定[13])。Bootstrap 检验(5000 次重复抽样, 参照 Preacher 与 Hayes 的间接效应检验策略[14])显示: 性别对能力目标达成度的总效应为 -0.070 ($P=0.002$); 纳入学习投入后, 直接效应降至 -0.049 ($P=0.041$), 间接效应为 -0.021 (95%CI: -0.048 — -0.004), 中介占比约 30.0%; 性别对学习投入的效应为 -0.52 ($P=0.006$)。按现有样本量计算, 该中介检验的效能约为 0.61, 故该结果仅作假设生成, 不足以支持因果推论。

3.7 敏感性分析

(1) 年级(时间)趋势检验。对照组三个年级理论闭卷成绩分别为 (77.9 ± 9.8) (78.6 ± 9.5) (79.6 ± 9.6) 分, 单因素方差分析显示年级间差异无统计学意义 ($F(2, 155)=0.58$, $P=0.560$, $\eta^2=0.007$), 提示不存在明显的“年级成熟”或“逐年提分”趋势, 削弱了以时间趋势解释组间差异的可能。(2) 权重稳健性。目标权重在 $\pm 10\%$ 范围内扰动后, 总目标达成度波动于 0.825—0.838 之间, 分档结论不变。(3) 社会赞许性调整。将 MCSDS-10 得分纳入模型后, 思政目标达成度由 0.86 降至 0.85, 仍为三类目标中最高。(4) 以最近一届(2023 级)为对照的再分析显示, 理论闭卷差值 2.76 分 ($P=0.021$)、模拟法庭辩论差值 7.94 分 ($P<0.001$), 与全样本结论方向一致、量级相当, 表明结论对对照组构成不敏感。

4. 讨论

4.1 主要发现

本研究有三点主要发现。第一, 在控制先修成绩、性别与年级后, 试验组在理论闭卷、合同审查报告与模拟法庭辩论三项共同测评上均优于历史对照, 其中模拟法庭辩论的效应量最大 ($d=1.05$), 提示以情境体验、角色扮演与价值辨析为特征的教学环节, 其增益大于以知识再现为特征的传统环节。第二, 课程总目标达成度为 0.83, 知识与思政目标进入“良好”档, 能力目标仍停留在“中等(达标)”档, 三者差异显著且效应量偏大 (偏 $\eta^2=0.150$), 说明多源证据模型能够敏感识别教学短板, 而不是给出笼统的“总体良好”。第三, 女生在能力与思政维度优于男生, 学习投入的解释力约三成。

4.2 理论贡献：TED 模型与三条可检验命题

本研究的理论增量不在于“把 OBE 与课程思政结合”——此类研究已不鲜见——而在于提供了可迁移的方法学构件与可检验的理论命题。

其一，提出“目标—证据—诊断”三链闭环模型（TED，图 1），把“思政要求→课程子目标→观测证据→达成度诊断”完整打通。其关键是思政目标的操作化：通过四维行为锚定量规与多主体、多情境的三角验证，把“法治信仰”“诚信守约”等抽象概念转化为可观察、可计分的行为指标，从而使思政成效进入达成度核算成为可能。

其二，厘清了达成度算法的学理归属。已有研究常把加权求和笼统称为“模糊综合评价”，却未交代隶属函数与合成算子。本文证明：在连续型评语集、线性隶属函数与加权平均型算子 $M(\cdot, +)$ 条件下，模糊综合评判解析地退化为加权得分率模型；这为使用加权得分率提供了明确的学理依据，也提示若采用离散评语集则须相应构造隶属函数与模糊矩阵，不可混用。

其三，基于本研究的发现提出三条可供后续研究检验的命题：（P1，情境敏感性命题）思政目标达成度对情境化、体验式考核环节权重的敏感性高于知识目标，即在其他条件不变时，情境化环节权重的提高会扩大思政与知识达成度的差值——本研究中思政（0.86）高于知识（0.84）远高于能力（0.78），为 P1 提供了初步证据，但需通过权重再配置实验加以检验；（P2，形式陌生性命题）新增表现性测评形式在首轮实施中存在“形式陌生抑制”，能力类达成度会随实施轮次递增而趋近其真实水平，因此首轮数据可能低估能力目标的真实达成度——这要求达成度评价至少跨两轮实施才能稳定比较；（P3，投入中介命题）性别等人口学变量对达成度的影响主要由学习投入与参与度中介，而非能力本身差异——本研究的探索性分析支持部分中介，但检验效能不足，需在更大样本、多课程情境下复验。三条命题使本研究从“一次课程改革”上升为可被后续研究证实或证伪的理论构念。

其四，确立“诊断而非排序”的达成度使用范式。本文把阈值与处置指向直接挂钩（表 4），使达成度结果自动指向“教学堵点—改进动作”，从而把评价嵌入持续改进链条，而非终结于一个分数。

4.3 因果推断的限度与混杂因素

必须坦承，本研究的试验组成绩更高不能直接归因于 OBE 改革。作为单校、单课程、非随机分组的历史对照研究，至少存在五类混杂：①年级与时间趋势——试验组为最后一届，可能受益于教学团队经验累积、题库与案例库成熟等，本文虽以逐年趋势检验与“最近一届对照”再分析加以缓解，但无法完全排除；②教师投入与期望效应（霍桑效应）——教师知晓改革意图，可能在试验组投入更多精力与情感；③测评工具扩展——对照组额外接受了合同审查与模拟法庭测评，这一安排虽提高了工具可比性，却可能同时抬升其动机水平，从而

低估真实组间差异，也可能因不熟悉表现性任务而拉低其得分，从而高估差异，方向不确定；④评分者盲法不完全——尽管隐去姓名与分组，评分教师仍可能通过答题风格与学期信息识别组别；⑤选择偏倚——受试来自单一院校单一专业，且对照组跨三届、试验组仅一届，两组在入学动机、班级氛围等方面的差异无法完全测量。

因此，本文对 RQ3 的回答是“关联”而非“因果”。后续研究宜采用整群随机或阶梯楔形（stepped-wedge）设计，在多校多课程情境下实施，并运用双重差分或倾向得分匹配等方法提高因果推断的稳健性；同时把教学团队、班级与学期作为随机效应纳入多层模型，以正确处理数据的嵌套结构。

4.4 思政目标测量效度的边界

思政目标达成度（0.86）最高，至少有三种竞争性解释，本文不作单一归因：①真实内化——情境体验确实有利于法治信仰与职业操守的内化；②评分宽松——教师对思政维度存在宽容偏差，尤其当同一教师同时承担教学与评价时；③工具不敏感——四维量规仅有 4 个等级，思政维度均分为 3.44/4，接近天花板，难以区分高分段学生。本文通过评分者培训、匿名同伴评定、社会赞许性协变量检验与行为编码三角验证部分缓解了这些问题，但 MCSDS-10 的检验效能受样本量限制，且同伴互评可能引入人情分。后续应引入情境判断测验（SJT）、延迟后测与实习单位行为评价作为效标，以检验思政达成度的预测效度。同样地，性别差异的解释亦须谨慎：本研究仅测量了学习投入，未测量学业动机、自我效能与学习策略，样本量（男 30 人）也不足以支撑稳定结论，故相关解释应视为假设而非结论。

4.5 实践启示

第一，把达成度用于诊断。若某一观测点出现群体性偏低，教师应锁定教学堵点，优化施教方式、充实案例库与教学资源；若偏低仅出现在个别学生身上，则应引导其追索原因、识别差距并主动补足弱项。第二，把能力目标作为优先改进项。具体举措包括：提前发布量规与评分样例，降低“形式陌生性”带来的抑制（对应 P2）；在首次实施情境化实训前增加一次不计分的演练轮；引入校企双导师，把真实合同审查任务引入课堂。第三，针对男生群体设计参与支架，如异质分组中的角色轮换、过程性即时反馈与阶段性目标设定。第四，建立评价方案的年度迭代机制，随行业实践与法律规范演进同步更新案例库、量规与权重，防止将评价凝固为一次性终端结论。

4.6 研究局限与展望

本研究的局限包括：①单校、单课程、单轮实施，外部效度有限；②非随机分组，因果推断受限（见 4.3）；③样本量有限，尤其在性别亚组与中介分析中检验效能不足；④仅有后测数据，缺少前测基线、延迟保持与行为迁移测量，无法判断成效的可持续性；⑤思政

测量工具的效标关联效度尚未建立；⑥未开展成本—效益分析，多维评价带来的教师工作量增加未予量化。后续研究可在三个方向推进：一是把 TED 模型推广至经管类其他核心课程与多所院校，检验其跨课程、跨校的稳定性与权重可迁移性；二是开展多轮次纵向追踪，检验 P2 所述“形式陌生性”随轮次递减的假设；三是结合毕业生跟踪调查与用人单位反馈，检验达成度对岗位合规行为的预测效度，从而完成从“学习证据”到“职业行为”的效度闭环。

5. 结论

本研究在高职专科金融服务与管理专业《经济法》课程中，构建并检验了“目标—证据—诊断”三链闭环（TED）达成度评价模型。研究表明：以专业培养目标和思政要求为起点逆向设计课程目标、以真实商业情境重构教学过程、以 7 个环节的多源证据同步考查知识、能力与思政三类目标，可把思政育人成效由“软指标”转化为可测量、可比较、可诊断的量化证据；试验组课程总目标达成度为 0.83，知识与思政目标进入“良好”档，能力目标为“中等”档，提示后续改进应聚焦于真实情境中的法律适用与风险识别训练。受单校、单课程与非随机分组限制，本文结论宜作关联性解释；TED 模型的可迁移性、权重稳定性与思政测量的效标效度，仍需在多校、多课程、多轮次的研究中进一步检验。

参考文献

- [1] 中华人民共和国教育部。高等学校课程思政建设指导纲要：教高〔2020〕3号[Z]. 2020-05-28.
- [2] 教育部新文科建设工作组。新文科建设宣言[Z]. 2020-11-03.
- [3] 中共中央办公厅，国务院办公厅。关于推动现代职业教育高质量发展的意见[Z]. 2021-10-12.
- [4] 李志义。解析工程教育专业认证的成果导向理念[J]. 中国高等教育, 2014(17): 7-10.
- [5] KILLEN R. Outcomes-based education: principles and possibilities[J]. Perspectives in Education, 2000, 18(1): 1-15.
- [6] SPADY W G. Outcome-based education: critical issues and answers[R]. Arlington: American Association of School Administrators, 1994.
- [7] HARDEN R M. Outcome-based education: the future is today[J]. Medical Teacher, 2007, 29(7): 625-629.
- [8] BIGGS J, TANG C. Teaching for quality learning at university: what the student does[M]. 4th ed. Maidenhead: Open University Press, 2011: 45-58.
- [9] 韩震。推进大中小学德育一体化进程的理念与思路[J]. 中国高等教育, 2020(17).
- [10] 谢季坚，刘承平。模糊数学方法及其应用[M]. 4版。武汉：华中科技大学出版社，2013.
- [11] SAATY T L. The Analytic Hierarchy Process[M]. New York: McGraw-Hill, 1980.
- [12] HASSON F, KEENEY S, MCKENNA H. Research guidelines for the Delphi survey technique[J]. Journal of Advanced Nursing, 2000, 32(4): 1008-1015.
- [13] FREDRICKS J A, BLUMENFELD P C, PARIS A H. School engagement: potential of the concept, state of the evidence[J]. Review of Educational Research, 2004, 74(1): 59-109.
- [14] HARDEN R M. AMEE Guide No. 21: curriculum mapping: a tool for transparent and authentic teaching and learning[J]. Medical Teacher, 2001, 23(2): 123-137.
- [15] 张守文。经济法原理[M]. 2版。北京：北京大学出版社，2020: 3-11.

Evaluating Course Objective Achievement Based on Professional Training and Ideological-Political Requirements under the OBE Concept: A Case Study of Economic Law in Higher Vocational Education

MA Yi, LI Chi-fang

(Foshan Polytechnic, Foshan, Guangdong 528137)

Abstract: [Objective] This study addresses the persistent challenges of "soft indicators being difficult to quantify" for ideological and political education effectiveness in economics and management courses, and the "overemphasis on knowledge while neglecting value" issue in curriculum goal achievement evaluation, so as to construct and validate a transferable achievement evaluation model. [Methods] A single-center non-randomized quasi-experimental study was conducted, with 158 junior college students from the Financial Services and Management Program (Cohorts 2021 - 2023) at Institution G serving as the historical control group, and 76 students from the 2024 cohort as the experimental group. A three-chain closed-loop "Target-Evidence-Diagnosis" (TED) model was established. The Delphi-Analytic Hierarchy Process (Delphi-AHP) was adopted to determine the weights of sub-curriculum objectives and 7 assessment links, and the weighted scoring rate method was used to synthesize the achievement degrees of three types of sub-objectives. Independent samples t-test, analysis of covariance (ANCOVA), repeated-measures ANOVA, and Bootstrap mediation effect test were applied for statistical inference, followed by sensitivity analysis and statistical power analysis. [Results] The experimental group scored significantly higher than the control group in the theoretical closed-book examination, contract review report, and moot court debate; the differences remained statistically significant after controlling for prior academic performance, gender, and cohort. In the experimental group, the overall curriculum goal achievement, knowledge goal, and ideological and political goal were rated as "Good", while the competence goal was rated as "Moderate", with statistically significant differences across the three dimensions. Female students outperformed male students in the competence and ideological-political dimensions, and learning engagement played a partial mediating role. [Conclusion] The TED model can transform the effectiveness of ideological and political education into measurable, comparable, and diagnosable quantitative evidence, providing a methodological path for the ideological and political evaluation of higher vocational economics and management courses. The competence objective and the male student subgroup are the key priorities for subsequent improvement. Limited by the constraints of a single institution, single course, and non-randomized grouping, the results of this study should be interpreted as correlational rather than causal.

Keywords: Outcome-Based Education (OBE); curriculum ideological and political education; curriculum goal achievement; Economic Law; higher vocational colleges; quasi-experimental study; fuzzy comprehensive evaluation