

机械安全技术课程目标达成度评价

许伟伟, 陈华俊, 贾小红, 孙嘉怡, 蔡之奕, 周厚昌

无锡科技职业学院智能制造学院, 江苏 无锡

收稿日期: 2026年8月5日; 录用日期: 2026年10月11日; 发布日期: 2026年10月20日

摘要

产出导向教育(OBE)要求课程目标达成度可量化、可审计,但是已有研究大多停留在报告达成度数值上,缺少对目标-试题映射完整性、试题测量质量、认知层次分布的联合诊断。以机械安全技术课程为案例,创建可复用的达成度评价基础层,并对其进行测量可信度的方法学检验。以83名高职学生A卷期末客观题(35题)为样本,建立两层目标体系(K1~K5、A1~A4、S1~S3)、Q矩阵、评分规则的评价框架,并且用经典测量理论来对试题质量进行诊断,即KR20与Cronbach's α 信度、项目难度 p 、点二列区分度 r 、按题型编码的布鲁姆认知层次分布。知识目标平均达成度为91.2%,可测能力目标(A1~A4)平均为91.7%,整体良好。试题质量诊断显示存在显著天花板效应(平均难度 $p = 0.916$)和普遍低区分度(23/35题, $|r| < 0.20$, $\alpha = 0.431$),说明达成度数值对区分学生能力的贡献很小。结论认为达成度基础层必须同时给出映射缺口和试题测量属性,才能作为TOPSIS等多准则决策的可靠数据基础层;仅依赖百分比不能用来推导出课程改进结论。

关键词

达成度评价, 产出导向教育, Q矩阵, 认知诊断

Evaluation of the Goal Achievement of Mechanical Safety Technology Course

Weiwei Xu, Huajun Chen, Xiaohong Jia, Jiayi Sun, Zhiyi Cai, Houchang Zhou

School of Intelligent Manufacturing, Wuxi Vocational College of Science and Technology, Wuxi Jiangsu

Received: August 5, 2026; accepted: October 11, 2026; published: October 20, 2026

Abstract

Outcome-based education (OBE) requires that the achievement of curriculum objectives can be quantified and auditable. However, most of the existing studies only focus on the value of report achievement, and lack of joint diagnosis of the completeness of target-question mapping, the quality

of question measurement and the distribution of cognitive level. Taking the course of mechanical safety technology as an example, a reusable basic layer of achievement evaluation is created, and the methodological test of measuring credibility is carried out. Taking 83 higher vocational students' A-volume final objective questions (35 questions) as samples, the evaluation framework of two-level target system (K1~K5, A1~A4, S1~S3), Q matrix and scoring rules was established, and the classical measurement theory was used to diagnose the quality of the questions, that is, KR20 and Cronbach's α reliability, item difficulty p , point two column discrimination r , Bloom's cognitive level distribution encoded by question type. The average achievement degree of knowledge goal is 91.2%, and the average measurable ability goal (A1~A4) is 91.7%, which is good as a whole. The diagnosis of test quality showed that there was a significant ceiling effect (average difficulty $p = 0.916$) and a general low discrimination (23/35 questions, $|r| < 0.20$, $\alpha = 0.431$), indicating that the achievement value contributed little to distinguishing students' ability. It is concluded that the reachability base layer must give the mapping gap and test measurement attributes at the same time, so as to be a reliable data base layer for multi-criteria decision-making such as TOPSIS. Pure percentage cannot be used to deduce the conclusion of curriculum improvement.

Keywords

Achievement Evaluation, Outcome-Based Education (OBE), Q-Matrix, Cognitive Diagnosis

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

产出导向教育(OBE)已经成为了工程教育专业认证的主要模式[1], 它的核心就是以可观测的学习产出反向设计课程目标, 并且用达成度评价检验教学闭环。当 OBE 理念和目标 - 前测 - 参与式学习 - 后测 - 总结的课堂结构相耦合时, 目标设定和评价环节就处在同一个闭环内互相校准, 高等教育情境下的实证研究显示这种耦合可以改善教学产出[2]。由此可见, 达成度评价不是课程末端的附加程序, 而是目标体系自我检验的一个部分。在高职安全技术与管理专业中, 机械安全技术对应着危险辨识、风险评估、防护设计的职业能力链条, 其达成度数据既是专业自评依据, 也是后续教学干预的关键依据[3]-[5]。既有相关研究的局限集中在三个方面, 即目标 - 试题映射多为人工勾画且缺乏完整性检验, 使部分目标无题可考却仍被计入而未加说明[6][7]; 达成度计算停留在百分比汇总层面, 鲜少报告试卷自身的信度与区分度, 使结论的测量前提不够明确[8]; 认知层次分布长期未被刻画, 难以判断评价是否覆盖记忆、理解、应用的不同深度[9]。

认知诊断模型(CDM)用 Q 矩阵来表示试题和属性之间的隶属关系, 已经成为技能诊断的主要工具[6]; Q 矩阵如果有标定错误, 那么模型参数估计就会出现误差, 并且属性估计和被试分类也会出错, 因此 Q 矩阵修正研究就致力于恢复正确的属性 - 项目标定[7]; 多值(序次) Q 矩阵设计进一步支持多级评分诊断[10]。达成度数值的可信度同样依赖试卷测量学属性: 经典测量理论对难度、区分度与内部一致性(库德 - 理查逊法、克伦巴赫系数法)指标的述评表明, 这些指标虽为试题质量评价的基础, 但其测算结果是否可信高度依赖考试实施与评分的客观性, 且难以孤立评判试卷质量[8]。Bloom 经典分类学及其修订版把认知过程分为记忆、理解、应用、分析、评价、创造六个层次, 是评价任务深度设计的基础标尺[9]; 建构一致性要求评价的认知层次与目标预期层次相匹。当客观题以记忆型判断为主、而应用型人才目标却依赖理解型单选支撑时, 层次错配会削弱达成度的解释力。将布鲁姆编码纳入达成度基础层, 正是为了显

式暴露此类错配。

既有研究或聚焦 OBE 框架与工程教育认证下的达成度评价设计[1] [3] [4]，或专注测评模型本身[6] [7] [10]，却较少将“目标映射完整性 + 试题测量质量 + 认知层次”三者整合于同一达成度基础层。职业教育所指向的复杂技能更凸显这一缺口：复杂问题解决过程中形成性评价的普遍缺位，既让学生难以自察薄弱，也使教师无从监测技能掌握的真实进程[11]，终结性试卷因而被赋予承担了它本不足以承担的诊断功能。基于证据强度的测量工具分级框架主张：分数含义取决于其使用与解释所依赖的测量前提[12]。安全工程教育格外强调能力可及性与情境化训练[3]-[5]。这三点恰是本研究将测量质量前置的理由。

本研究以机械安全技术课程为例，采用 83 名高职学生 A 卷期末客观题成绩，建立两层目标体系(K1~K5 知识、A1~A4 能力、S1~S3 素养)，并用 Q 矩阵显式表达目标 - 试题映射；进而引入经典测量理论的信度、难度、区分度与布鲁姆认知层次诊断，对达成度数值背后的测量质量做实证检验。拟回答两个问题：第一，达成度评价的基础层应如何系统化构建，并使目标映射可审计；第二，当试卷存在天花板效应与低区分度时，达成度结论应受何种约束。为聚焦测量方法学，不展开子群体层面的细分比较。

2. 研究方法

2.1. 数据来源与样本

样本为本校安全技术与管理专业 83 名二年级学生的 A 卷期末成绩。达成度量化以客观题为对象：单选 20 题、多选 5 题、判断 10 题，共 35 题、满分 65 分。

2.2. 三层目标体系与集合建模

将课程目标建模为三层互不相交的集合之并：

$$G = K \cup A \cup S, |K| = 5, |A| = 4, |S| = 3 \quad (1)$$

其中 K 为知识目标(K1 安全法规与标准、K2 机械危险源辨识、K3 通用与专用防护、K4 防护装置与装备、K5 全生命周期管理)， A 为能力目标(A1 危险识别、A2 风险评价、A3 防护选择、A4 标准查阅)， S 为素养目标(S1 安全意识、S2 规范习惯、S3 责任担当)。

2.3. Q 矩阵与目标 - 试题映射

设客观题集合 Q 含 35 题，目标集合 G 共 12 项(K1~K5、A1~A4、S1~S3)，其中客观题可测者 9 项，定义 Q 矩阵：

$$Q = (Q_{q,g}) \in \{0,1\}^{35 \times 12} \quad (2)$$

对每个目标计算覆盖题数 $c_g = \sum_q Q_{q,g}$ ，其中 $Q_{q,g}$ 表示第 q 题考查目标 g 。若 $c_g = 0$ ，则该目标在客观题中无支撑，达成度记为不可计算。表 1 给出完整的试题 - 目标 Q 矩阵；各可测目标覆盖题数为：K1 = 9、K2 = 10、K3 = 8、K4 = 6、K5 = 2、A1 = 20、A2 = 6、A3 = 7、A4 = 2；S1~S3 的 $c_g = 0$ 。

Table 1. Q-matrix

表 1. Q 矩阵

题号	题型	满分	K1	K2	K3	K4	K5	A1	A2	A3	A4
Q1	单选	2	√					√			
Q2	单选	2	√					√			
Q3	单选	2		√				√			

续表

Q4	单选	2		√			√	
Q5	单选	2		√			√	
Q6	单选	2		√			√	
Q7	单选	2	√				√	
Q8	单选	2	√					√
Q9	单选	2		√			√	
Q10	单选	2		√			√	
Q11	单选	2		√			√	
Q12	单选	2	√					√
Q13	单选	2			√		√	
Q14	单选	2			√		√	
Q15	单选	2			√			√
Q16	单选	2			√			√
Q17	单选	2				√		√
Q18	单选	2			√			√
Q19	单选	2			√			√
Q20	单选	2					√	√
Q21	多选	3	√				√	
Q22	多选	3		√			√	
Q23	多选	3				√		√
Q24	多选	3				√		√
Q25	多选	3			√			√
Q26	判断	1	√				√	
Q27	判断	1		√			√	
Q28	判断	1	√				√	
Q29	判断	1	√				√	
Q30	判断	1		√			√	
Q31	判断	1				√		√
Q32	判断	1			√			√
Q33	判断	1			√			√
Q34	判断	1			√			√
Q35	判断	1					√	√

2.4. 评分规则

对第 q 题、第 i 名学生, 得分 $s_{q,i}$ 取下列值:

$$s_{q,i} = \begin{cases} M_q, & \text{与标准答案一致} \\ 0, & \text{单选/判断错选} \\ M_q, & \text{多选全对} \\ m_q, & \text{多选少选(部分给分)} \\ 0, & \text{多选错选} \end{cases} \quad (3)$$

其中 M_q 为第 q 题满分, m_q 为多选少选的部分。该规则解释了为何客观题整体得分偏高——答对即得满分、少选部分给分, 使得分分布偏向高端。

2.5. 达成度与变异系数

目标 g 的达成度 A_g 与知识点 p 的达成度 A_p 定义为实际得分和相对满分之比:

$$A_g = \frac{\sum_i \sum_{q:Q_{q,g}=1} s_{q,i}}{\sum_i \sum_{q:Q_{q,g}=1} M_q} \times 100\% \quad (4)$$

$$A_p = \frac{\sum_i s_{p,i}}{\sum_i M_p} \times 100\% \quad (5)$$

为刻画群体分化, 计算变异系数 $CV_g = \sigma_g / \mu_g$ (σ_g 、 μ_g 分别为该目标得分率的标准差与均值)。CV 越大, 学生间掌握程度越不均衡。

2.6. 试题质量模型

用经典测量理论指标来评价试卷质量。内部一致性用严格二分题(得分只有 0 或满分)计算 KR₂₀, 用全部 35 题计算 Cronbach's α :

$$\alpha = \frac{Q}{Q-1} \left(1 - \frac{\sum_p \sigma_p^2}{\sigma_X^2} \right) \quad (6)$$

$$KR_{20} = \frac{J}{J-1} \left(1 - \frac{\sum_j p_j (1-p_j)}{\sigma_{X_j}^2} \right) \quad (7)$$

项目难度用通过率 $p_q = \bar{s}_q / M_q$ 估计, 值越接近 1 越易, 其中 $\bar{s}_q$ 为第 q 题的被试平均得分; 区分度用题目得分与总分的点二列相关 r_q 估计, $|r_q| < 0.20$ 视作区分力不足。经典的试题质量述评认为, 区分度是题目对不同水平被试心理特质的区分能力, 其计算结果同样需要得分信息的可靠性, 信度(如库德-理查逊 KR₂₀、Cronbach's)的误用也应当被重视。布鲁姆认知层次按题型规则编码, 判断为 L1 记忆, 单选为 L2 理解, 多选为 L3 应用, 据此统计各层次题量分布, 检验评价对各个认知层次的覆盖广度。

3. 结果

3.1. 两层目标达成度

表 2 给出 9 个可测目标的达成度与等级。知识目标平均 91.2%, 可测能力目标(A1~A4)平均 91.7%, 整体落入良好至优秀区间。K1 安全法规与标准(95.5%)、K5 全生命周期管理(96.8%)与 A4 标准查阅能力(96.8%)居于高位; K4 防护装置与装备(84.6%)与 A3 防护选择能力(85.4%)相对靠后, 但仍高于合格线。

图 1 展现了上述梯度。

Table 2. Achievement and grade levels of the two-tier objectives
表 2. 两层目标体系目标达成度与等级

目标	名称	达成度(%)	等级
K1	安全法规与标准	95.5	优秀
K2	机械危险源辨识	89.7	良好
K3	通用与专用防护	89.3	良好
K4	防护装置与装备	84.6	良好
K5	全生命周期管理	96.8	优秀
A1	危险识别能力	91.0	优秀
A2	风险评价能力	93.5	优秀
A3	防护选择能力	85.4	良好
A4	标准查阅能力	96.8	优秀

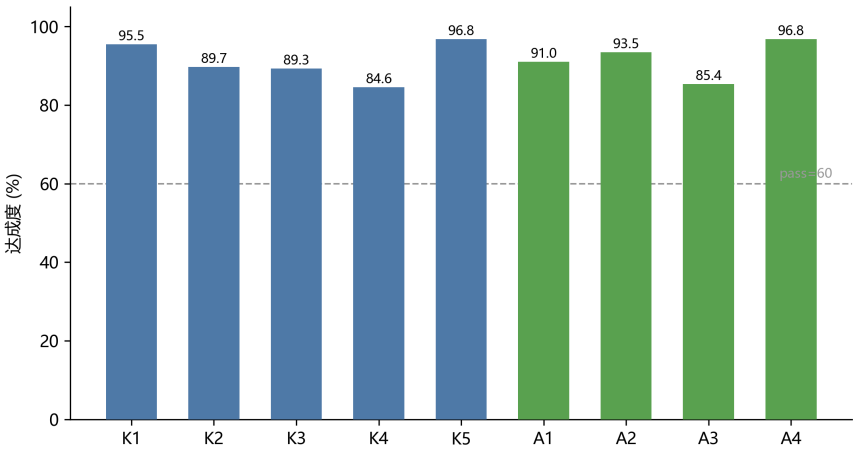


Figure 1. Two-tier objective achievement
图 1. 两层目标达成度

3.2. 知识点达成度与薄弱环节

Table 3. Knowledge-point achievement and coefficient of variation
表 3. 知识点达成度与变异系数

编号	知识点	模块	达成度(%)	CV
KP03	安全防护装置原理	M3	69.0	0.670
KP15	职业卫生基础	M4	72.2	0.587
KP09	防护装置类别	M3	74.6	0.434
KP04	机械安全防护设计	M3	81.7	0.336
KP13	个体防护装备	M3	83.3	0.248
KP11	危险源管理	M2	86.5	0.215
KP06	起重/特种设备安全	M5	90.1	0.212
KP08	人机工程学基础	M3	90.5	0.324

续表

KP10	安全操作规程	M2	91.9	0.199
KP07	金属切削机床安全	M6	94.9	0.124
KP14	安全标志与警示	M3	95.6	0.159
KP16	机械安全标准	M1	96.4	0.192
KP17	安全文化建设	M1	96.8	0.147
KP01	机械安全基础概念	M1	97.6	0.075
KP12	事故应急救援	M4	98.0	0.109
KP05	电气安全基础	M4	98.4	0.103
KP02	机械危险辨识	M1	98.6	0.075

依据客观题实际考查的旧知识点，连续重编号为 KP01~KP17。表 3 按达成度升序列出全部 17 项，最薄弱者为 KP03 安全防护装置原理 69.0% (CV 0.670)、KP15 职业卫生基础 72.2% (CV 0.587)与 KP09 防护装置类别 74.6% (CV 0.434)，前两项集中于防护设计模块 M3 与电气安全 M4，提示该知识簇存在系统性薄弱。图 2 以升序条形给出 17 个知识点的达成度(括号内为 CV)与分化程度：分布明显右偏，除底部三项外多高于 85%，整体掌握良好；薄弱点却非随机分布，而呈簇状出现在防护设计相关模块。这三项的 CV 均越过 0.43，意味着均值偏低之外，学生间掌握程度也高度分化：有人几近未掌握，有人接近满分。教学干预因此应瞄准低均值、高分化簇，而非面向全体补差。

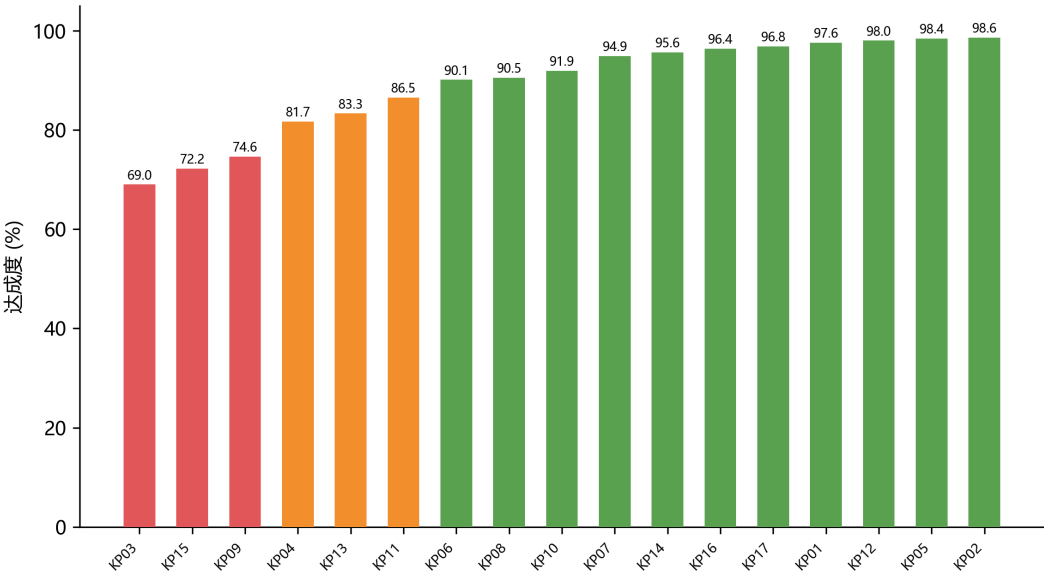


Figure 2. Knowledge-point achievement
图 2. 知识点达成度

3.3. 题型得分率

表 4 给出三类客观题的得分率。单选 94.8%、判断 92.5%居高位，多选仅 77.1%，构成相对短板。这与多选全对得满分、少选部分得分、错选零分的阶梯式赋分结构一致，反映学生在综合辨识(应用层次)上更吃力。图 3 对比三类题型可知，得分率与布鲁姆认知层次呈单调对应：判断(L1 记忆)92.5%最高，单选

(L2 理解) 94.8%次之，多选(L3 应用)骤降至 77.1%，落差约 17 个百分点。这一梯度恰由赋分结构放大造成，即多选错选零分、少选部分给分扩大了应用层的综合辨识难度。学生不是不会判断，只是不能在很多正确的要素中做出完整的选择。应用能力才是客观题评价的主要短板。

Table 4. Score rates by objective-item type
表 4. 客观题型得分率

题型	题数	满分	得分率(%)
单选	20	40	94.8
多选	5	15	77.1
判断	10	10	92.5

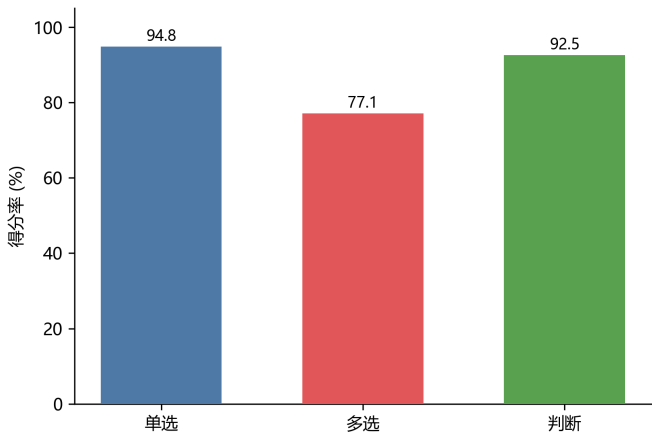


Figure 3. Comparison of scoring rates for objective question types
图 3. 客观题型得分率对比

3.4. 试题质量与项目分析

表 5 汇总测量学指标。全卷 Cronbach’s $\alpha = 0.431$ (偏低)，而 30 道严格二分题的 $KR_{20} = 0.798$ (可接受)。平均难度 $p = 0.916$ ，呈强天花板效应；平均区分度 $|r| = 0.134$ ，且 23/35 题 $|r| < 0.20$ 、最高 $|r|$ 仅 0.427，说明大量题目难以区分高低能力学生。布鲁姆层次分布为 L1 记忆 10 题、L2 理解 20 题、L3 应用 5 题，应用层次占比偏低，与目标中能力权重不匹配。图 4 以难度 - 区分度散点显示多数题目落入易且低区分象限，虚线为 $|r| = 0.20$ 警示线。

Table 5. Summary of item quality
表 5. 试题质量汇总

指标	值	说明
Cronbach’s α (全卷 35 题)	0.431	内部一致性偏低，受天花板效应与少区分题拖累
KR-20 (30 道严格二分题)	0.798	30 道二分题，可接受
平均难度 p	0.916	强天花板效应，整体偏易
平均区分度 $ r $	0.134	r 为题得分与总分的点二列相关，均值低，整体区分力弱
低区分题数 ($ r < 0.20$)	23/35	近三分之二题目区分不足，难以分辨高低能力学生
布鲁姆分布 L1/L2/L3	10/20/5	应用层次(L3)题量偏少，与能力目标权重不匹配

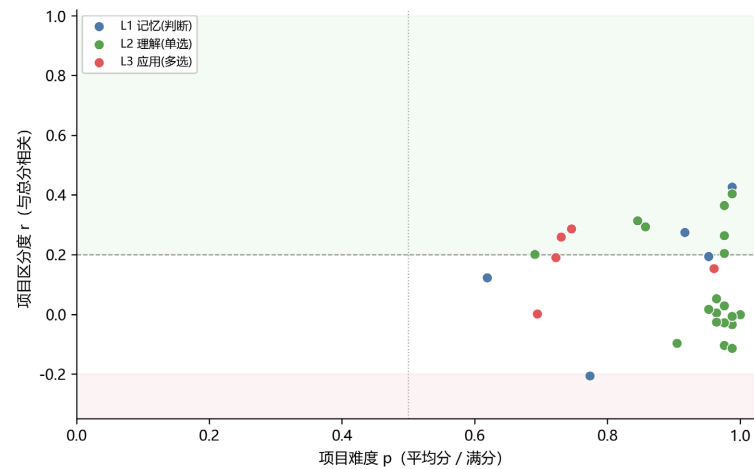


Figure 4. Item difficulty-discrimination scatter
图 4. 试题难度 - 区分度散点

模块层面(表 6、图 5)的聚合结果进一步收敛到同一结论：防护设计模块 M3 达成度最低(82.4%)且 CV 最高(0.362)，是分化最显著的薄弱区；基础模块 M1 (97.4%)与设备安全 M6 (94.9%)最为稳固。薄弱并非均匀分布，而是锁定在防护装置原理、类别与设计这一知识簇。

Table 6. Module achievement aggregation
表 6. 模块达成度聚合

模块	名称	知识点数	平均达成度(%)	平均 CV
M1	Basis	4	97.4	0.122
M2	Hazard ID	2	89.2	0.207
M3	Protection	6	82.4	0.362
M4	Electrical	3	89.5	0.266
M5	Special Equip	1	90.1	0.212
M6	Machine Equip	1	94.9	0.124

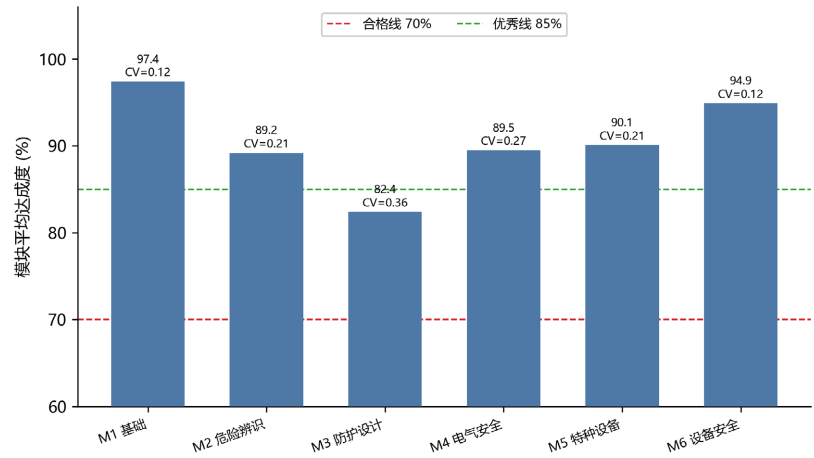


Figure 5. Module achievement (M1~M6)
图 5. 模块达成度(M1~M6)

4. 讨论

4.1. 映射完整性决定达成度的可解释性

Q 矩阵审计暴露出三个知识点的映射缺口。若沿用凡列目标皆计达成度的惯例,不可计算项将被混入排名,进而误导课程改进方向。本研究将缺口显式标注为 N/E,正是对基于证据强度的测量工具分级框架与 Q 矩阵映射完整性原则的呼应:分数使用须以测量前提为根基,达成度数值唯有在映射完整时才具备解释力。

4.2. 试题质量对达成度结论的约束

研究表明达成度数值的解释须受试题测量属性的约束。本卷呈现显著的天花板效应(平均难度 $p=0.916$):高通过率意味着试卷在能力量表的上端所能提供的信息十分有限,整体优秀在很大程度上反映度量上限被压缩,而非学生能力普遍达标的充分证据。区分度分析显示 23/35 题的 $|r|<0.20$,且全卷 Cronbach's α 仅为 0.431,表明试卷对不同水平被试的区分效力有限。在此前提下,若仅依据达成度数值在个体或知识点层面做精细排序,其结论的稳健性难以保证。因此,达成度基础层必须同步报告难度、区分度与信度等测量属性——唯有将达成度与支撑它的测量质量一并呈现,才能区分某知识点达成度最低究竟源于真实的能力偏差,还是源于题目质量缺陷本身。需要明确的是,二者不可混同:本卷 α 仅为 0.431 且近三分之二题目区分度不足,意味着达成度数值更多刻画的是测量工具自身的局限,而非对学生能力的无偏估计;对其解释应限定在“试卷所呈现的结果”层面,避免据之对个体能力做出过高或过低的因果推断。

4.3. 基础层作为多准则决策的审计基础

上述成果(目标映射、知识点达成度、测量质量指标)可以直接用于 TOPSIS、熵权法等决策方法构造决策矩阵——用知识点达成度做方案评分, Q 矩阵覆盖数和区分度作权重修正项,使得排序不但能回答谁最低,也能回答这个值是否可以采信。

4.4. 局限与改进方向

本研究所存在的不足之处有:一是只选取了某一门课和一份试卷作为样本,得出的结果不能进行广泛的外推;后续应纳入多课程、多学年乃至多院校数据复现结论;二是没有对主观题进行量化,素养目标 S1~S3 还是无法得到直接测量;今后应将主观题以分级评分者信度(ICC)与内容分析法纳入分析框架三是横截面的数据缺少跨学期的复测,重测信度还不可知;四是区分度受到天花板效应的压制,需要更具有区分力的题目来重新构建试卷。五是布鲁姆认知层次由题型规则推断,存在“题型 = 认知层次”的简化风险;应邀请至少两名学科专家按布鲁姆分类学定义独立编码每题,并以 Cohen's Kappa 系数检验编码一致性,对信度极低的题目先修订试卷、待测量质量合格后再行达成度分析。基于以上问题,在今后的工作中可以从三个方面展开。首先是在评价模型上可以考虑使用模糊认知诊断来融合并存的主观表现,用认知诊断模型估计属性掌握的概率代替目前的二分达成度,提高基础层信息量及区分精度。其次是从教学组织的角度来看,对于应用层的相对短板,应该把练习序列由单主题集中训练改为跨主题交错编排,利用交错练习在保持与迁移的优势加强防护设计知识簇的掌握。最后是在评价机制上要以覆盖学习全过程的形成性任务补足终结性试卷的诊断局限,通过“目标-前测-后测”闭环的设计使得改进措施的效果能够被再次测量而不是仅仅凭借经验做出判断。受样本规模限制采用经典测量理论(CTT)而非项目反应理论(IRT)或认知诊断模型(CDM):IRT 参数估计需更大样本方能稳定, $n=83$ 不足以支撑其标定;后续可扩大样本后引入 IRT 估计能力潜变量,并借 CDM [6] [7] 实现知识点的诊断,与本框架形成方法学对话

与互补。

5. 结论

本文以机械安全技术课程为例,构建了“两层目标体系 + Q 矩阵 + 评分规则”达成度评价的基础层,并把经典的测量理论中的信度、难度、区分度以及布鲁姆的认知层次诊断融入其中。实证结果显示:知识目标平均达成度 91.2%,可测的能力目标平均 91.7%,总体在良好到优秀的水平上(但受本卷低区分度与天花板效应制约,该数值反映试卷呈现的表面达成度,不能视为对学生真实能力的无偏估计),但是有 3 个知识点由于映射缺口不能进行计算;薄弱点锁定在防护设计模块 M3 (KP03/KP09/KP15) (须谨慎:因 23/35 题区分度不足,该知识点层级定位应视为初步线索,尚不足以单独支撑教学干预),而且多选题型构成了相对短板;强烈的强天花板效应和低区分度提醒本文,达成度数值必须与测量质量一同出现才能成为课程改进及多准则决策的可审计依据。这一基础层不仅可以给一线老师提供诊断教学问题、定位薄弱知识簇的实证依据,也为后续的 TOPSIS 等多准则决策的研究提供了透明、可复现的数据基础层。

基金项目

无锡市哲学社会科学招标课题(高等教育专项)(市级):人工智能赋能高职课程培养目标达成度评价研究——以机械安全技术课程为例, WWSK26-GJ-C09。

参考文献

- [1] 闵力,阮娟. 基于 OBE 理念的高校课程目标达成度评价体系构建与实践[J]. 创新教育研究, 2026, 14(3): 53-60.
- [2] Xu, Z.W., Ge, L.P., He, W., et al. (2024) Developing a BOPPPS (Bridge-In, Objectives, Pre-Assessment, Participatory Learning, Post-Assessment and Summary) Model Combined with the OBE (outcome Based Education) Concept to Improve the Teaching Outcomes of Higher Education. *Humanities and Social Sciences Communications*, 11, Article No. 1001. <https://doi.org/10.1057/s41599-024-03519-y>
- [3] 曾艳,肖莉,聂礼飞. 工程教育专业认证背景下防火防爆技术课程目标达成度评价[J]. 科教导刊, 2026(11): 32-35.
- [4] 王晋,田晓,季淮君,等. 基于工程教育认证的课程目标达成度评价实践——以系统安全工程课程为例[J]. 安全, 2024, 45(9): 92-96.
- [5] 安金鹏,王焕. 高职高专安全技术管理专业机械安全课程存在问题及改革思路[J]. 内燃机与配件, 2017(7): 141-142.
- [6] 宋丽红,汪文义,丁树良. 认知诊断评估中 Q 矩阵理论及应用[J]. 心理科学进展, 2024, 32(6): 1010-1030.
- [7] 刘彦楼,吴琼琼. 认知诊断模型 Q 矩阵修正: 完整信息矩阵的作用[J]. 心理学报, 2023, 55(1): 142-158.
- [8] 仲红. 经典教育测量理论试题质量评价指标述评[J]. 教育实践与研究, 2008(9): 7-9.
- [9] 王毅力,伦小秀,洪喻. 修订版布鲁姆教育目标分类学的环境化学教学设计[J]. 高等理科教育, 2026(1): 23-32.
- [10] 蒋婷婷,宋丽红,汪文义,等. 在不同计分方式下的多值 Q 矩阵设计理论[J]. 江西师范大学学报(自然科学版), 2025, 49(5): 450-456.
- [11] Xu, X.L., Shen, W.Q., Islam, A.Y.M.A. and Zhou, Y. (2023) A Whole Learning Process-Oriented Formative Assessment Framework to Cultivate Complex Skills. *Humanities and Social Sciences Communications*, 10, Article No. 653. <https://doi.org/10.1057/s41599-023-02200-0>
- [12] 刘雨今,华雨婷,黄维肖,等. 基于证据强度的测量工具分级框架简介[J]. 循证护理, 2023, 9(16): 2919-2924.