

基于优序层次分析的学生形成性评价指标体系构建

程肖冰¹, 王晓蕾^{2*}, 杨世元^{1*}

¹北京联合大学城市轨道交通与物流学院, 北京

²北京联合大学教务处, 北京

收稿日期: 2025年4月25日; 录用日期: 2025年5月23日; 发布日期: 2025年5月30日

摘要

目前的学生形成性评价内容片面, 存在评价不够客观、全面的问题, 为解决这些问题, 本研究通过问卷调查获取数据, 并基于数据相关有效性分析, 层次分析法初步确定二级指标权重分配, 优序图确定指标体系权重分配, 最终构建学生形成性评价指标体系。结果表明, 基于优序层次分析的学生形成性评价指标体系不仅能够帮助教师更准确地了解学生的学习状态和进步情况, 还能够为学生提供有针对性的学习建议和反馈, 促进他们的全面发展。

关键词

形成性评价, 层次分析法, 优序图法

Construction of Student Formation Evaluation Index System Based on Hierarchical Priority Analysis

Xiaobing Cheng¹, Xiaolei Wang^{2*}, Shiyuan Yang^{1*}

¹College of Urban Rail Transit and Logistics, Beijing Union University, Beijing

²Academic Affairs Office, Beijing Union University, Beijing

Received: Apr. 25th, 2025; accepted: May 23rd, 2025; published: May 30th, 2025

Abstract

The current formative assessment of students is one-sided, with issues of insufficient objectivity

*通讯作者。

文章引用: 程肖冰, 王晓蕾, 杨世元. 基于优序层次分析的学生形成性评价指标体系构建[J]. 教育进展, 2025, 15(5): 1612-1621. DOI: 10.12677/ae.2025.155947

and comprehensiveness. To address these problems, this study collected data through a questionnaire survey and, based on relevant validity analysis of the data, used the Analytic Hierarchy Process to preliminarily determine the weight distribution of secondary indicators. The priority graph was used to determine the weight distribution of the indicator system, ultimately constructing a formative assessment indicator system for students. The results show that the formative assessment indicator system based on priority-level analysis not only helps teachers more accurately understand students' learning status and progress but also provides targeted learning suggestions and feedback to students, promoting their all-round development.

Keywords

Formative Evaluation, Analytic Hierarchy Process (AHP), Priority Matrix Method

Copyright © 2025 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

教育评价引导高校学生树立学习目标, 促进能力提升, 激励学生自我评价[1]。作为教育评价的重要一环, 形成性评价通过持续反馈帮助学生及时调整学习策略, 强化自主学习能力, 引导其关注过程改进而非单一结果, 促进知识建构与能力发展, 同时培养元认知意识与批判性思维, 对高校学生的学术成长具有动态导向作用。我国学者已开始对基于人工智能、大数据等新兴技术构建形成性指标体系的方法展开探索, 近3年的研究数量明显增加。

丁婷等通过文献综述、实证研究、案例分析等方法, 分析了教育数据挖掘在识别学生特征、预测就业趋势、优化培养方案等方面的作用, 并提出了建议[1]。孟凡等集成到现有的教育应用中, 在保护隐私的基础上使用联合可视化框架可以最大限度地提高模型的准确性, 联合学习将为教育的信息化和智能化发展提供一条新的途径[2]。赵海霞等开展了创新创业教育分层教学研究与实践, 数据驱动的学员分层可以精准地实现创新创业教育分层教学, 更好地满足学员个性化的学习需求, 提升他们的创新创业能力[3]。吴龙凯等构建应以“人脑”基本结构和“类脑”功能体系等研究为基础, 依托数据、算法、算力三个核心要素, 实现数据挖掘、传输、集成、分析、呈现、应用全过程闭环, 展现未来教育评价数据大脑的落地应用[4]。牛义锋等精准应用于课程思政教育, 有助于深入挖掘学生个性化需求, 进而为高校大学生思想政治教育的精准化供给提供有力支撑[5]。高艺璐等基于 Stacking 集成学习的方法, 构建组合模型对大学生的学业成绩进行预测, 提高学习资源分配效率并有效实施个性化教学[6]。

目前现有研究存在: 评价体系构建方式方法单一, 构建依据缺乏对数据的系统性, 综合性支撑和学生形成性评价内容片面, 存在评价不够客观、全面的问题。2020年, 中共国务院印发的《深化新时代教育评价改革总体方案》就明确指出: 要完善评价结果运用, 综合发挥导向、鉴定、诊断、调控和改进作用。现有研究为本文教育实践提供了丰富的理论指导和实践参考, 一是运用层次序法构建指标体系, 为指标选择的层面提供系统性的方案; 其次, 以数据挖掘作为依据, 精确捕捉学习行为数据, 明确且精准建立评价标准。本研究基于以上思想, 将数据信度效度检验、层次分析法、优序图分配权重有机地结合起来, 形成优序层次分析法。此外, 本研究创新性地将问卷调查与数据挖掘技术相结合, 系统地设计评价指标, 确保评价维度的全面性和明确性, 为形成性评价提供客观、精确的数据支持。这种有机结合不仅能够增强评价体系的科学性和有效性, 还能够推动形成性评价在教育实践中的广泛应用与深入发展。

2. 理论基础

2.1. 形成性评价指标体系

形成性评价(Formative Assessment)是教学过程中的一种评价方式,其目的在于提升学生的学习成效,并推动教育教学的持续改进。与传统的总结性评价不同,形成性评价侧重于对学习过程的持续跟踪、监控和反馈,以便教师能够及时调整教学方法,从而更有效地促进学生的学习。

基于数学模型构建的形成性评价的应用可以实现教育评价精准化、多元化。通过对评价终端学生的学习轨迹数据分析,实现数字化追踪学习者的知识掌握轨迹,自动识别认知盲区并生成动态学情图谱。根据目前已有可行的数学模型,本文根据数学模型价值链分析(见表 1),构建优序层次分析法,运用组合评价法的系统性整合优势,确定作为构建指标体系的方法,实现从单一模型片面分析到多模型系统分析的跨越。

Table 1. Mathematical model value chain analysis
表 1. 数学模型价值链分析

| 方法 | 适用场景 | 优势 | 问题 |
|---------|---------------------------|--------------------|--------------------|
| 层次分析法 | 适用于多准则决策,用于确定评价指标的权重分配 | 定性与定量结合,减少主观偏差 | 权重计算容易偏差,需分层归类 |
| 灰色综合评价法 | 适用于数据量少、信息不完全的评价场景 | 无需数据归一化,可直接使用原始数据 | 评价结果稳定性不足 |
| 模糊综合评价法 | 处理难以量化的评价指标,通过定性评价转化为定量分析 | 支持多层级评价,适用于复杂多维度评价 | 需要避免主观权重分配导致的偏差 |
| 优序图法 | 适用于确定多维度评价指标的优先级或权重 | 操作简单直观,动态适应,减少主观偏差 | 指标数量限制,无法处理细化指标 |
| 组合评价法 | 整合多种模型,解决单一模型的局限性 | 解决单一模型局限性 | 需要通过整合多种模型,提升评价鲁棒性 |

2.2. 学习进阶过程

学习进阶过程划分是多层次组合分析方法的前置依据,想要实现多层次赋权的基础就是根据学习进阶划分,确定各阶段指标权重确定方法。在划分管理学课程之前,首先需要根据学生的学习认知阶段和教学实际,明确学生的学习进阶过程。现有高校学生学习进阶分为:课前预习、课堂学习、课后巩固三个阶段。其中具有管理学课程鲜明特点的学习内容包括三大部分:课前学习有线上平台预习与课前导入案例学习两大内容;课堂学习包括课堂知识点导入、课堂知识点讲解、课堂知识点扩展三大内容;课后巩固包括课后延展案例学习、课后阶段测试和课后延展学习三大内容。

其中,需要特别注意的是,在完成预习阶段后,进入到课程学习阶段,主导学习的角色会有所转换,教师会作为教学阶段的主要角色,需要通过相应的知识点或背景案例,导入课堂授课知识点的主要授课内容。随后对课程的各个知识点进行讲解,并对知识点进行讲解。

进入到课后巩固阶段,主要角色回到学生,学生需要完成知识点的阶段测试,巩固所学内容。完成知识点巩固后,根据延展案例,提升学生对知识点的应用能力。教学流程划分见图 1。

2.3. 优序层次分析法

优序层次分析法的本质是优序图法与层次分析法的组合分析法,有机结合优序图法的是指基于层次分析和优序图赋权有机结合的数据分析方法。主要包括三部分内容:数据相关与有效性验证,初步挖掘输入模型数据,确保其适应性;层次分析法初步确定权重,通过层次分析法对不同授课阶段的指标赋权,完成二级指标体系构建;最后利用优序图法完成指标体系构建。

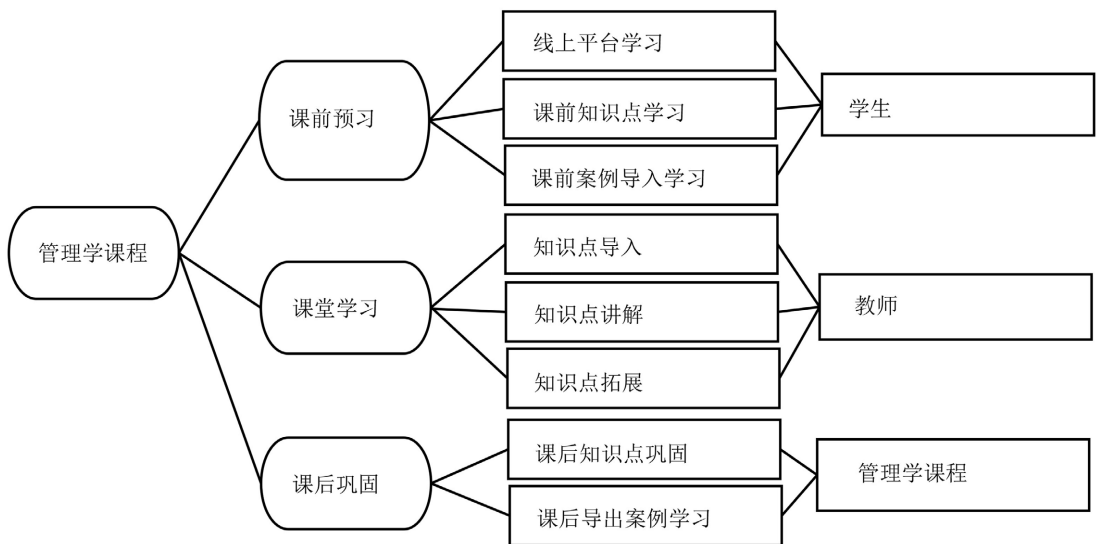


Figure 1. Division of teaching progression process
图 1. 教学进阶流程划分

2.3.1. 数据有效相关性检验

优序层次分析法构建指标体系的本质是：对一组相互之间具有强关联性的数据进行逐步的数据挖掘，核心在于输入数据中各变量之间是否相关，数据本身是否具有有效性。

本文验证数据相关性及其有效性采取的是 KMO 检验，其优势为 KMO 检验的优势在于其客观量化、明确标准、变量筛选指导性以及对偏相关的敏感性，辅助高效判断数据质量并排除不合理研究项。KMO 检验的结果通常以数值的形式呈现。如果 KMO 指数的值在 0.6 到 1.0 之间，则认为数据适合做因子分析其计算模型如式(1)所示。

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} \alpha_{ij}^2} \tag{1}$$

式中 r_{ij} 表示简单相关系数， $\alpha_{ij,1,2,3,\dots,k}^2$ 表示偏相关系数。

2.3.2. 层次分析法确定二级指标权重

层次分析法是一种定性与定量相结合的决策分析方法。它是一种将决策者对复杂系统的决策思维过程模型化、数量化的过程。层次分析确定二级指标权重主要有三个步骤：

- (1) 建立层次结构，根据课程阶段数据进行分组，把每个课程阶段设计的指标作为一个层次，按照目标层、若干中间层(准则层)以及最低层(方案层)的形式排列起来。
- (2) 构造判断矩阵，设有 n 个指标： $\{A_1, A_2, A_3, \dots, A_n\}$ ， a_{ij} 表示 A_i 相对于 A_j 的重要程度判断值。 a_{ij} 一般取 1, 3, 5, 7, 9 等 5 个等级标度，判断矩阵 A 如式 2 所示：

$$A = \begin{bmatrix} \frac{w_1}{w_1} & \dots & \frac{w_1}{w_n} \\ \frac{w_1}{w_1} & & \frac{w_n}{w_n} \\ \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \dots & \frac{w_n}{w_n} \end{bmatrix} \tag{2}$$

层次排序。层次单排序的目的是对于上层次中的某元素而言，确定本层次与之有联系的元素重要性的次序。它是本层次所有元素对上一层次而言的重要性排序的基础。若取权重向量 $W = [w_1, w_2, \dots, w_n]^T$ 成

立，则有：

$$A \times W = \lambda \times W$$

(3)

λ 是 A 的最大正特征值，那么 W 是 A 的对应于 λ 的特征向量。从而层次单排序转化为求解判断矩阵的最大特征值 λ_{\max} 和它所对应的特征向量，就可以得出这一组指标的相对权重。为了检验判断矩阵的一致性，需要计算它的一致性指标：

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

(4)

当 $CI = 0$ 时，判断矩阵具有完全一致性；反之， CI 愈大，则判断矩阵的一致性就愈差。
为了检验判断矩阵是否具有令人满意的一致性，则需要将 CI 与平均随机一致性指标 RI (见表 2) 进行比较。一般而言，1 或 2 阶判断矩阵总是具有完全一致性的。

Table 2. Average random consistency index RI
表 2. 平均随机一致性指标 RI

| 阶数 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|------|------|------|------|------|------|------|
| RI | 0 | 0 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1.41 | 1.45 | 1.49 | 1.52 | 1.54 | 1.56 | 1.58 | 1.59 |

对于 2 阶以上的判断矩阵，其一致性指标 CI 与同阶的平均随机一致性指标 RI 之比，称为判断矩阵的随机一致性比例， CR 如式 5 所示时：

$$CR = \frac{CI}{RI} \leq 0.10$$

(5)

我们就认为判断矩阵具有令人满意的一致性；否则，当 $CR \geq 0.10$ 时，就需要调整判断矩阵，直到满意为止。

2.3.3. 优序图法确定指标体系

数据有效优序图法是穆迪(P. E. Moody)在 1983 年提出的方法，针对需要比较的 n 个比较对象(例如设计方案、目标、价格指标等)分别进行排序，并分别对各个序号进行评分，即得到优序数，然后进行综合评价，分别计算各个评价指标的总优序数，按照总优序数数值大小进行评定。

Table 3. Steps of determining index system by priority graph
表 3. 优序图法确定指标体系步骤

| 步骤 | 步骤内容 | 核心目标 |
|-------|--------------------------------------|--------------------|
| Step1 | 评价者对比各指标进行评价 | 专家评价法确定评价体系构建依据 |
| Step2 | 针对各个对比目标所在的行进行相加，得到该目标的总优序数 | 确定指标优序，为优序总和计算提供基础 |
| Step3 | 综合计算所有对比目标的总优序数，得到优序数总和 | 确定总优序，为指标权重赋权提供依据 |
| Step4 | 各个对比目标的总优序数除以优序数总和， 获得该对比目标的权重百分比 | 确定评价指标权重 |
| Step5 | 根据各指标权重百分比形成指标体系 | 形成指标体系 |

分析过程为：评价者针对具体的评估方案进行排序，两两对比建立判定矩阵，用 0 和 1 来表示方案的优劣：1 表示两两指标对比中，相对更加重要的；0 则代表不重要的；若两者对比相当，则可以采用评分 0.5。优序图法确定指标体系步骤详见表 3。

3. 学生形成性评价指标体系形成

3.1. 数据有效相关性检验

为了便于对问卷结果进行数据计算分析，问卷主体结构分为一级指标(A)、二级指标(B1~B4)、三级指标(C1~C6)、四级指标(D1~D15)、意见选项五个部分。其中，意见选项设置为“非常合适”、“合适”、“一般”、“不合适”、“非常不合适”五个，量化处理分别赋值 5、4、3、2、1。2024 年 5 月 8 日到 2023 年 9 月 1 日，组织完学生的问卷调查。共发放问卷 1000 份，收回问卷 922 份，问卷有效度 92.2%。

3.2. 指标权重体系形成

3.2.1. 数据有效相关性检验

信度检验和效度检验采用 KMO 和巴特利特检验法，用来检验评估一组数据是否呈现出一定程度的相关性和可信性。由表 4 可知：KMO 值的权重是 0.887，非常适用提取分析数据，表明通过了信度检验；Bartlett 的球形度检验卡方值的权重是 3784.992，适合进行因子分析；显著性的权重是 0.000，且单因子标准化载荷均大于 0.5，表明调查数据具有较好的效度，具体结果见表 4。

Table 4. Data reliability analysis
表 4. 数据信度分析

| KMO 和 Bartlett 的检验 | | | |
|--------------------|------|----------|---------------------|
| 方法 | | 取值 | 分析 |
| KMO 值 | | 0.887 | 大于 0.8，非常适用提取分析 |
| Bartlett 球形度检验 | 近似卡方 | 3784.992 | 适合进行因子分析 |
| | df | 136 | 自由度，无意义 |
| | p 值 | 0.000 | 通过 Bartlett 检验，具有效度 |

3.2.2. 基于层次分析法确定二级指标权重

| 序号 | 课前学习 | 课堂学习 | 课后学习 | |
|--------|-------|-------|-------|-------|
| 85 | 29 | 59 | 12 | |
| 86 | 35 | 37 | 28 | |
| 87 | 35 | 40 | 25 | |
| 88 | 34 | 52 | 14 | |
| 89 | 10 | 60 | 30 | |
| 平均值 | 28.6 | 49.6 | 21.8 | |
| 构建判断矩阵 | 1 | 0.509 | 1.019 | |
| | 1.963 | 1 | 2.001 | |
| | 0.981 | 0.5 | 1 | |
| 求和 | 3.944 | 2.009 | 4.02 | 行求和 |
| 归一化 | 0.253 | 0.253 | 0.253 | 0.759 |
| | 0.498 | 0.498 | 0.498 | 1.494 |
| | 0.249 | 0.249 | 0.249 | 0.747 |
| 权重 | 0.254 | 0.498 | 0.249 | |
| 最大特征值 | 3 | RI | 0.52 | |
| CI | 0 | CR | 0 | |

Figure 2. Steps for calculating the weights corresponding to study time periods
图 2. 学习时间段对应权重计算步骤

由上文可知，问卷数据通过了相关性及有效性检验，证明适用于优序层次分析法。故运用层次分析确定二级指标权重(学习阶段指标权重)，其计算步骤见图 2。

从图 2 可知，针对课前学习、课堂学习、课后学习总共 3 项，构建 3 阶判断矩阵进行和积法层析分析，分析得到特征向量为：0.759, 1.494, 0.747, 3 项学习阶段对应的权重值分别是：25.4%, 49.8%, 24.9%。

本次针对 3 阶判断矩阵计算得到 CI 值为 0.000，针对 RI 值查表为 0.52，因此计算得到 CR 值为 0.000 < 0.1，意味着本次研究判断矩阵满足一致性检验，计算所得权重具有一致性。进一步利用 Matlab 进行检验，计算结果见图 3。

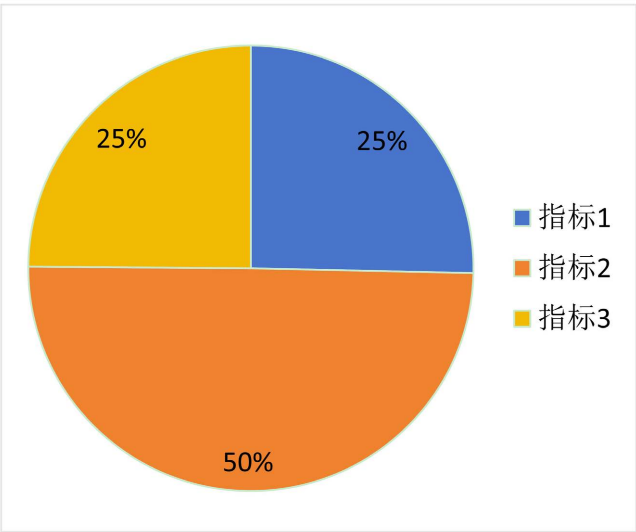


Figure 3. Weight validation of indicators corresponding to study time periods
图 3. 学习时间段对应指标权重检验

3.2.3. 基于优序图法确定指标体系权重

基于优序图分析，对 15 个指标进行排序，并分别对各个序号进行评分，即得到优序数，然后进行综合评价，分别计算各个评价指标的总优序数，按照总优序数数值大小进行评定，基于此构建指标权重体系。本文利用 Matlab 进行检验并完成指标权重，计算结果见图 4~6。

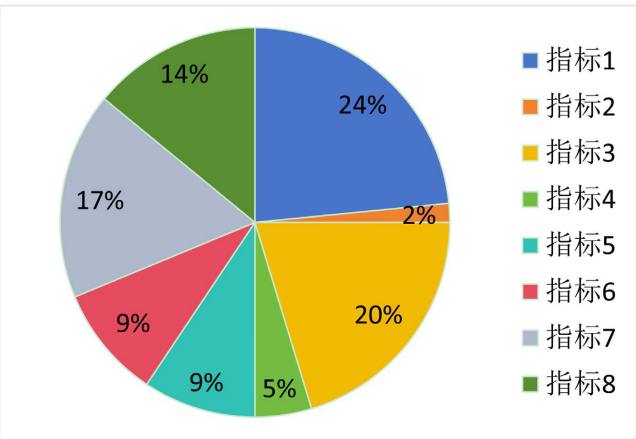


Figure 4. Pre-class learning weight calculation test
图 4. 课前学习权重计算检验

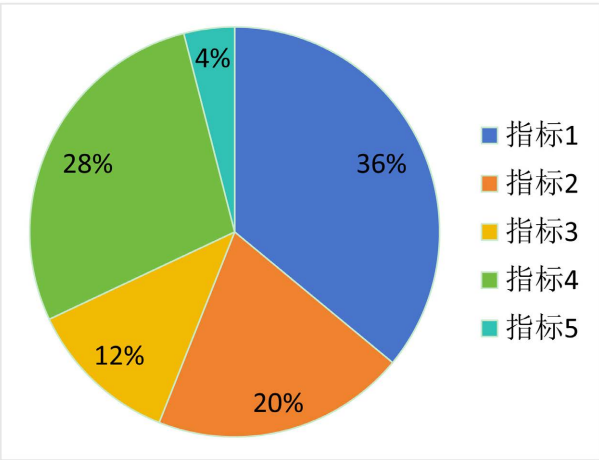


Figure 5. Test of weight calculation for classroom learning
图 5. 课堂学习权重计算检验

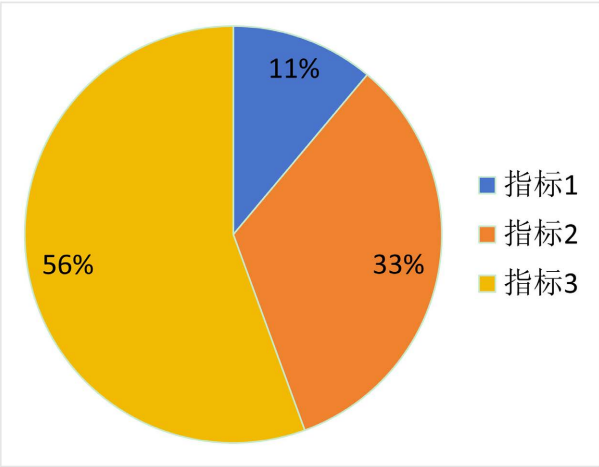


Figure 6. Post-class and independent study weight calculation test
图 6. 课后与自主学习权重计算检验

3.3. 学生形成性评价指标体系确定

基于数据相关有效性分析，层次分析法初步确定二级指标权重分配，优序图确定指标体系权重分配，其构建的学生形成性评价指标体系及权重见表 5。

Table 5. Student formative evaluation index system
表 5. 学生形成性评价指标体系

| 一级指标 | 一级指标权重(%) | 二级指标 | 二级指标权重(%) | 最终权重(%) |
|------|-----------|------------------|-----------|---------|
| 课前学习 | 25.35 | 自主学习 教学平台的出勤率 | 23.44 | 5.94 |
| | | 平台观看教学资源的次数 | 1.56 | 0.40 |
| | | 平台学习进度是否按照老师要求进行 | 20.31 | 5.15 |
| | | 记录自学笔记的情况 | 4.69 | 1.19 |
| | | 查阅相关前沿资料(如文献)的情况 | 9.38 | 2.38 |

续表

| | | | | | |
|------|-------|------|---|-------|-------|
| 课堂学习 | 49.77 | 自主学习 | 平台作业完成的情况: (1) 从学生开始答题到提交答案的时长(10~50 分钟) | 9.38 | 2.38 |
| | | | (2) 作业正确率 | | |
| | | | 平台自学测验的正确率 | 17.19 | 4.36 |
| | | | 协作学习 小组成员共同完成小组作业 | 14.06 | 3.56 |
| | | | 课堂出勤次数 | 36.00 | 17.92 |
| | | | 课堂学习的氛围(富有热情、思维活跃, 表情自然, 语言清晰、举止大方) | 20.00 | 9.96 |
| | | 协作学习 | 课上回答问题的次数及准确性: 案例分析时, 思路清晰、条理清楚, 能运用 PPT 展示案例, 效果美观 | 12.00 | 5.97 |
| | | | 提出的问题有一定的深度或具有创新性 | 28.00 | 13.94 |
| | | | 组内的交流提问(寻求帮助或解答)及问题汇报(提出问题或解决问题) | 4.00 | 1.99 |
| | | 巩固学习 | 完成老师布置的作业: (1) 从学生开始答题到提交答案的时长(10~50 分钟) | 11.11 | 2.76 |
| | | | (2) 作业正确率 | | |
| | | | 低成就者将本次课程仍然不会的问题及时询问老师或同学, 高成就者参加相应学科竞赛并获得不同程度奖项 | 33.33 | 8.29 |
| | | | 思政教育 学生课下了解案例, 自己实现 PPT 的制作, 表现出积极价值观, 实现与老师的双向奔赴 | 55.56 | 13.82 |

根据如上评价体系, 进行建模, 得到量化的打分, 见式(6)。

$$\gamma = \sum \sum \alpha_{ij} \chi_{ij} \tag{6}$$

γ 为学生总分, i 为第 i 个学习时间段, $i = 1, 2, 3$, j 为第 j 个二级指标, α_{ij} 为第 i 个学习时间段的第 j 个二级指标的权重, χ_{ij} 为教师在第 i 个学习时间段的第 j 个二级指标的对应分数。

4. 结论

本研究基于数据相关有效性分析, 层次分析法初步确定二级指标权重分配, 优序图确定指标体系权重, 成功构建了基于数据挖掘的学生形成性评价指标体系。该体系涵盖了课前学习、课堂学习和课后学习三个主要方面, 共包括一级指标 6 项、二级指标 16 项。该体系不仅关注学生的学习状态和进步情况, 强调发挥学生的优势。在学习进阶过程中, 通过不断评价, 及时提供有针对性的学习建议和反馈。此外, 通过量化分析, 确定了各级指标的权重, 提高了评价的准确性和可靠性, 并探索将竞赛融入评价体系, 可进一步形成学生评价流程库, 以持续优化和完善学生形成性评价体系。基于本文的研究路径及结果, 成功运用优序法确定减少指标权重赋权的主观偏差, 并运用层次分析实现定性与定量结合的指标体系细化赋权, 最终形成的组合评价法实现了从单一模型片面分析到多种方法相互补足, 各取所长的系统性分析的创新, 为进一步实现教育评价合理化, 数字支撑化, 步骤轻量化、结果实时化、评价体系可量化方向发展, 为构建下一代智能教育评价体系的核心方法论提供了新的研究路径与方法。

基金项目

北京联合大学 2023 年教育教学研究与改革项目(JJ2023Y033); 北京市教委科技一般资助项目(KM202011417002)。

参考文献

[1] 丁婷, 龚祯, 朱辰杰. 数字经济背景下基于教育数据挖掘的大学生高质量就业培养路径研究[J]. 商业经济,

2025(2): 191-193.

- [2] 孟凡, 海涛, 张瑞华, 等. 基于视觉联合学习的教育数据挖掘隐私保护技术[J]. 电子器件, 2023, 46(6): 1661-1672.
- [3] 赵海霞. 基于数据挖掘的创新创业教育分层教学研究与实践[J]. 广东开放大学学报, 2024, 33(1): 12-18.
- [4] 吴龙凯, 张珊, 刘姚慧卓, 等. 教育评价数据大脑的内涵、模型与路径[J]. 广西师范大学学报(哲学社会科学版), 2024, 60(6): 87-97.
- [5] 牛义锋, 叶珺, 徐秀珍. 基于大数据画像的精准化课程思政实现路径研究[J]. 中国教育信息化, 2024, 30(11): 95-102.
- [6] 高艺璐, 刘鸿雁, 姚明海, 等. 基于 Stacking 集成学习的学生成绩预测方法[J]. 电子设计工程, 2024, 32(24): 63-67.