

项目化学习的表现性评价设计研究——以人工智能课程为例

Research on the Design of Performance Assessment in Project - based Learning: Taking the Artificial Intelligence Course as an Example

文桥^{1*}, 沈诗淼²

¹² 江南大学 江苏“互联网+教育”研究基地

*1456678609@qq.com

【摘要】中小学人工智能课程中的项目化学习目前面临着缺乏与之匹配的评价方法或评价体系,本研究聚焦初中人工智能项目化学习表现性评价,通过构建“目标-任务-评价标准”设计框架与实施路径,为人工智能课程的项目化学习提出一套可操作的表现性评价体系和评价方案;最后,在某初中开展两轮教学实践验证。在人工智能课程项目化学习中开展表现性评价,有助于推动教学评一体化协同增质,增强学生学习人工智能的自主性与参与度,促进学生高阶能力的形成,更好地推进人工智能教育。

【关键词】人工智能课程;项目化学习;人工智能素养;表现性评价

Abstract: At present, the project-based learning in AI courses in primary and secondary schools is facing a lack of matching evaluation methods or evaluation systems. This study focuses on the performance evaluation of AI project-based learning in junior high schools. By constructing the design framework and implementation path of "goal task evaluation standard", this paper proposes a set of operable performance evaluation system and evaluation scheme for the project-based learning in AI courses; Finally, two rounds of teaching practice were carried out in a junior high school. The performance evaluation in the project-based learning of artificial intelligence course is helpful to promote the integration of teaching evaluation and collaborative quality improvement, enhance students' autonomy and participation in learning artificial intelligence, promote the formation of students' high-level ability, and better promote artificial intelligence education.

Keywords: artificial intelligence course; project-based learning; AI literacy; performance assessment

1.问题的提出

项目化学习(PBL)凭借高情境性、高参与度和问题导向的优势,成为中小学人工智能课程开展的主要方法(Biggs & Collis, 2014; Brennan & Resnick, 2012),其评价需通过多元手段采集学生真实情境中的综合表现,兼顾过程与成果(Cai & Merlino, 2011)。然而实践中常因评价理念缺失、教师缺乏科学设计方法,导致评价与教学脱节,难以实现教学评一体化

(Dagienė & Futschek, 2008; Chang & Lee, 2010),亟需构建兼顾学生项目表现与成果、整合评价与教学的结构化评价体系。而表现性评价以其评价目标的复合性、评价形式的一致性、评价实施的可行性、评价融合的统一性(Grant & Jay, 2017; Khattri & Sweet, 1996),为项目化学习的课堂评价提供了良好的思路。但其实施面临双重困境:教师对评价设计与流程认知模糊(Kong, Chiu, & Lai, 2018),且实际操作中易流于形式化,表现为过度关注成果、评价主体单一及反馈延迟(Korkmaz et al., 2017; Marzano et al., 1993)。本研究基于“目标-任务-标准”框架,结合教学案例探索表现性评价的设计与应用路径,为解决当前人工智能课程项目化学习中的表现性评价应用难题提供可行参考方案。

2.项目化学习的表现性评价设计框架

项目是开展素养导向下人工智能课堂教学的基本单位,而贯穿项目化学习的评价任务是表现性评价嵌入人工智能课堂中项目化学习的关键作用点,表现目标-表现任务-表现标准是

表现性评价嵌入项目化学习的关键路径。项目化学习的表现性评价设计应该包括目标-任务-标准三个基本要素，其设计框架如图 1 所示：

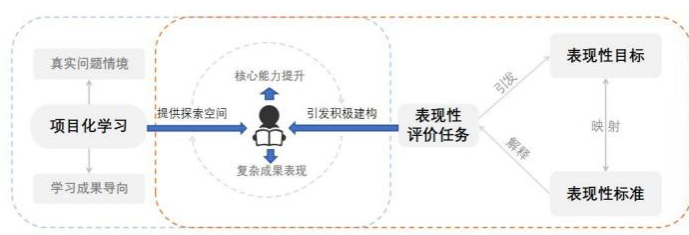


图 1 面向项目化学习的表现性评价框架

表现性评价目标是项目化学习评价设计的起点，有效的表现性目标应涵盖知识基础、能力发展和情感态度三重维度（起点水平—素养建构—人文温度），与课堂教学目标、教学内容、课程整体素养框架相衔接；表现性评价任务是表现性评价设计的核心，其设计既需要理清在什么项目进程设计符合什么特征的评价任务，还需要设计评价任务的呈现方式和数据观测点；评价标准需要建立目标维度/指标——任务执行表现——层级水平之间的映射关系，可借用量规（Rubric）当作具体罗列评价维度与评分标准的评分工具（陈素平，2020）。这需自下而上梳理项目化学习目标的必备维度和表现性评价任务的实施情况，形成具有明确的、基于进阶的层级水平描述，为教学评一体化提供结构化参照。

3. 基于人工智能课程的项目化学习表现性评价设计与实施

3.1. 面向人工智能课程 PBL 的表现性评价体系

本研究选取初中人工智能课堂作为研究场景，基于第 2 章构建的理论框架，并按照“目标-任务-标准”流程设计人工智能项目化学习评价体系（图 2）：以项目化学习阶段中问题的提出、问题的解决、成果的呈现为评价时间线，以情境分析、问题凝练、问题理解、建构方案、项目作品整理、项目展示交流等为关键评价任务设计并实践具体评价体系内容。通过课堂观察、作品量表等多模态工具采集数据，设置反思环节实现教学评闭环，既贯穿“知识→能力→情感”三维框架，又通过成果公开与修订环节推动学生高阶思维发展，整体凸显过程性、实践性与发展性特征。

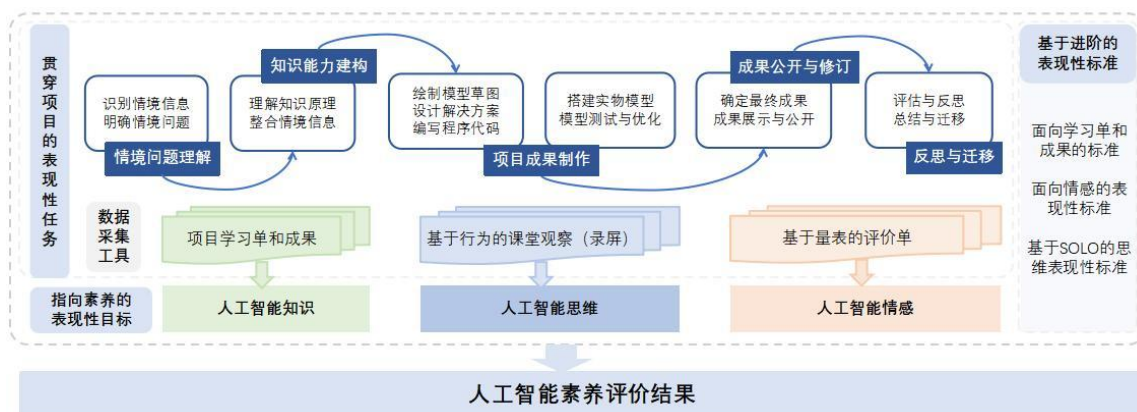


图 2 面向人工智能 PBL 的表现性评价体系图

3.2. 实验准备

本研究在华东地区某初中开展教学实验，选取初一年级学生 45 名（男女 1:1）人工智能素养薄弱但积极性高的学生，使用优必选 uKitExplore 机器人（基于 Arduino）和 Ucode4.0 图形化编程平台。通过预设性数据（项目记录单、作品评分表）与自然性数据（PC 端录屏）采集知识、思维水平表现，并结合师生问卷评估评价有效性。课程前两周进行设备熟悉与基础编程训练，后六周实施项目化学习，重点观察学生在问题解决高阶思维发展。

3.3. 两轮评价实施

第一轮试验基于“智能温室”项目，采用自上而下目标设计法，确立人工智能知识、情感、思维三维评价目标，第一轮实验具体内容如表 1 所示。首轮评价实施因任务描述模糊、试题难度失衡导致数据无效（如学生仅机械应答“知道/不知道”），且存在学生互评主观偏差，评价标准层级不清等问题。

表 1 第一轮表现性评价任务的详细内容

项目内容	描述
核心任务	模型搭建：绘制智能温室草图并组装传感器、主板等硬件 代码编写：实现温湿度监测与报警功能（串口输出、灯光控制）
数据收集	学生编程行为录屏记录、学习单记录过程性表现、组内互评（分工协作）、组间互评（实用性/创新性）评价情感维度。
实施流程	情境探索：分析限制条件，完成问题学习单；方案设计：小组协作设计外观、搭建模型并编写代码；调试优化：试运行程序、修正故障并填写检测报告；总结迁移：展示成果，自评/互评后完善项目单。

第二轮行动研究选择的任务主题是“智能机械臂”，并对第一轮存在的问题进行改进：使用清晰、具体的语言描述任务目标，增加形容词（如“准确地”“完整地”）和程度副词分层；补充情境线索，拆分复杂任务并提供支架（如将画流程图改为填关键步骤）；强制差异打分（小组成员得分须不同），标记无效分数（如全 0/全满分）；细化标准层级，采用等级制（合格/优秀）替代分值，增强引导性。

3.4. 数据分析与问题反思

针对两轮实验，基于学生编程行为的录屏数据，采用编码框架追踪学生操作序列；结合学习单内容分析、项目成果评分及多主体评分（自评 30%、互评 40%、师评 30%）；通过互评量表量化学生兴趣度与认知负荷水平。在实施效果验证中，第二轮专家评分均值显著提升（ICC>0.5），表明改进后设计更具可靠性，结果如表 2 所示：

表 2 第一轮与第二轮实施效果对比分析结果

两位专家在高质量 PA 量表的评分情况		
	均值	一致性（ICC）
第一轮	3.7	0.526（单个）
		0.690（均值）
第二轮	4.5	0.600（单个）
		0.750（均值）

另外，本研究还对两轮实验中学生的兴趣度和认知负荷水平以及学生编程行为的录屏数据进行分析，受篇幅所限，具体图表分析从略，但数据分析结果仍支持以下结论：评价设计与学生表现具有强关联性，第二轮实验通过调整评价任务和评价标准，能够显著提升评价效度与学生参与度。最后，评价设计需选取贴近学生日常，任务知识范畴不宜过小，数据采集工具设计为学生提供充足的表现空间，报告设计宜以开放性问答为主。评价设计过程要尽量规避问题情境和学生报告给学生分析和决策的引导和暗示，必要时辅以案例规则讲解。

4. 结语

本研究构建了中小学人工智能 PBL 表现性评价体系，以人工智能素养为核心目标，通过明确评价任务、标准及工具，为课堂实践提供可操作的解决方案。研究表明，优化后的评价

设计显著提升学生兴趣度并降低认知负荷。未来可拓展至多学段应用研究，并结合图像识别、文本分析等技术开发自动化评价系统，进一步提升基础教育中表现性评价的便捷性与实效性。

参考文献

- 陈素平. (2020). 让评价支持和促进学习——美国 STEM 项目学习评价的特点与启示. 上海教育科研, (4), 18 - 22.
- Grant, W., & Jay, M. (2017). 追求理解的教学设计 (2nd ed.). Translated by Yan, H. B., Song, X. L., Lai, P., et al. 上海: 华东师范大学出版社.
- Biggs, J. B., & Collis, K. F. (2014). Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). Academic Press.
- Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In Proceedings of the 2012 annual meeting of the American educational research association (pp. 25). Vancouver, Canada.
- Cai, J., & Merlino, F. J. (2011). Metaphor: A powerful means for assessing students' mathematical disposition. In D. Brahier & W. Speer (Eds.), Motivation and disposition: Pathways to learning mathematics (pp. 147 - 156). National Council of Teachers of Mathematics 2011 Yearbook. Reston, VA: NCTM.
- Chang, L. C., & Lee, G. C. (2010). A team-teaching model for practicing project-based learning in high school: Collaboration between computer and subject teachers. Computers & Education, 55(3), 961 - 969.
- Dagienė, V., & Futschek, G. (2008). Bebras international contest on informatics and computer literacy: Criteria for good tasks. In International conference on informatics in secondary schools—evolution and perspectives (pp. 19 - 30). Springer, Berlin, Heidelberg.
- Khatti, N., & Sweet, D. (1996). Assessment reform: Promises and challenges. In Implementing performance assessment: Promises, problems, and challenges (pp. 1 - 21).
- Kong, S. C., Chiu, M. M., & Lai, M. (2018). A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. Computers & Education, 127, 178 - 189.
- Korkmaz, O., Cakir, R., & Ozden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). Computers in Human Behavior, 72, 558 - 569.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). Assessing student outcomes: Performance assessment using the dimensions of learning model. Association for Supervision and Curriculum Development.