

基于课堂观察的 STEM 教学评价：协议、要素与方法

詹泽慧 吕思源

(华南师范大学教育信息技术学院, 广东广州 510631)

【摘要】 区别于传统讲授式课堂, STEM 课堂具有学科融合性、真实情境性、项目承载性, 然而如何对其展开评价一直是个难题。作为循证评价的重要方式, 课堂观察可为 STEM 教育质量评估提供可靠的教学实践证据与衡量标准。本文通过分析和比较国际典型的六套 STEM 课堂观察协议(RTOP、OTOP、UTOP、TDOP、COPUS、PORTAAL)发现: 在构成要素上, 各协议均体现了基于科学探究、先验知识、沟通合作、学科融合的教学方法, 发展学生概念思维、批判思维、创新思维与元认知能力的认知目标, 倡导自主、民主、互动的课堂文化氛围; 在应用情境上, 虽然所选协议能够从教学策略、课堂动态、师生互动等方面改善教学, 但普遍存在观测学生学习行为不足、忽视技术作用、难以表征实践操作等问题。据此, 本文基于探究共同体理论和 TPACK 框架将 STEM 课堂解构为教学存在、社会存在、技术存在三个维度, 统摄 STEM 课堂观察的教师行为与学生行为, 以学生为中心构建 STEM 课堂观察的 L-PST 双罗盘模型; 基于评价设计、证据收集、评价实施三环节构建 STEM 课堂观察评价的三维矩阵, 为教师和研究人員理解 STEM 课堂、选取适切的观察视角与评价工具、实施 STEM 课堂教学评价提供思路和借鉴。

【关键词】 STEM 教育; 教学评价; 课堂观察; 教师专业发展

【中图分类号】 G423.04 **【文献标识码】** A **【文章编号】** 1007-2179(2023)04-0065-16

一、引言

STEM 教育作为跨学科学习的典型范式, 是发展学生核心素养的有效手段(刘晟等, 2016)。与传统分科教学相比, STEM 教育具有学科融合性(Bybee, 2013)、真实情境性(English, 2016)、项目承载性(Erdogan et al., 2016)等特征, 聚焦实际问题, 有机整合多学科的知识与技能形成综合性问题解决方案(余胜泉等, 2015), 将原本孤立、分散的抽象知识与学生的生活经验相联系(秦瑾若等, 2017),

将知识蕴含于情境化的真实问题中(张屹等, 2017), 以学生为中心形成连贯有组织的课程结构, 在协作探究与具身体验中促进学生主动学习。

然而, STEM 教学评价是学界公认的难题(宋乃庆等, 2022; 胡艺龄等, 2022)。传统的纸笔测试不适用于素养导向的 STEM 教学, 强调从证据中获取信息的循证教学评价成为国际 STEM 教育实践的基本遵循(Rosenberg et al., 2018; 余胜泉等, 2019), 为我国 STEM 课堂教学改革提供了思路。课堂观察能够提供直接可靠的教学实践证据与衡

【收稿日期】 2023-05-24 **【修回日期】** 2023-06-20 **【DOI 编码】** 10.13966/j.cnki.kfjyyj.2023.04.007

【基金项目】 国家自然科学基金面上项目“基于事理图谱的计算思维智能导训模型及可解释性研究”(62277018); 教育部人文社科基金“基于 C-STEAM 的粤港澳大湾区教育协同创新机制研究”(22YJC880106); 华南师范大学哲学社会科学重大培育项目“面向创新能力培养的跨学科组合策略与应用效果研究”(ZDPY2208)。

【作者简介】 詹泽慧, 博士, 教授, 博士生导师, 青年珠江学者, 华南师范大学教育信息技术学院, 研究方向: STEAM 教育、学习分析、智慧教育、创新创业教育(zhanzehui@m.scnu.edu.cn); 吕思源, 硕士研究生, 华南师范大学教育信息技术学院, 研究方向: STEAM 教育。

【引用信息】 詹泽慧, 吕思源(2023). 基于课堂观察的 STEM 教学评价: 协议、要素与方法[J]. 开放教育研究, 29(4): 65-80.

量标准, 在 STEM 教学评估中备受青睐(Ebert-May et al., 2011)。

目前代表性的课堂观察工具集中在智慧教育领域, 包括基于信息技术的互动分析编码系统(ITLAS)、改进型弗兰德互动分析系统 iFIAS、基于教学活动的课堂教学行为系统(TBAS), 但这类工具存在观察视角平面化和片面化的局限, 难以反映学科特征、课型特点与教学风格(许娜等, 2020), 无法从整体上还原 STEM 课堂教学全过程。当前我国 STEM 教育评价以引介国际评估框架为主, 存在两方面问题: 1) 理论研究缺乏对课堂观察工具的比较分析, 给研究者们选择和借鉴带来困难; 2) 实践研究停留在本土化框架构建与自我验证阶段, 使用范式仍需从国际大规模评估项目中吸收经验。鉴于此, 本文以国际典型的 STEM 课堂观察协议为研究对象, 对其构成要素与适用情境进行内容分析, 总结已有工具的优势与不足, 形成 STEM 课堂观察的基本思路, 为我国 STEM 课堂教学评价的设计与实施提供参考。

二、研究方法

(一) 选择依据

有效的 STEM 课堂观察协议应该体现教与学的理论特征, 包括有效教学、课堂动态、教学方法、师生互动等细节(Ebert-May et al., 2011), 进而评估

具体教学措施实施的准确性与有效性; 实践上需要提供可靠的数据与有意义的反馈, 供教师理解并改善教学(Lund et al., 2015)。基于此, 本文选择六套标准化课堂观察协议进行比较分析(见表 1), 所选协议需满足三个条件: 1) 针对 STEM 课堂教学质量评估; 2) 具备详尽的协议研发文献; 3) 经过大规模实证数据验证。表 1 列举了协议的基本信息和选择依据。根据数据记录类型的不同, 观察协议可分为整体性协议(RTOP、OTOP、UTOP)和分段式协议(TDOP、COPUS、PORTAAL)两类(AAAS, 2012)。整体性协议针对课堂教学全过程开展整体性评价, 分段式协议要求观察员分时段评估每个观察项目。

(二) 分析框架

为揭示协议的构成要素和实践特征, 本文从协议构建的有效性、与 STEM 课堂教学的适切性、评价的适用性从三方面分析上述六套协议, 分析框架见表 2: 1) 构建层。参照美国国家科学研究委员会(National Research Council, NRC)提出的评价三角模型, 从认知、观察、解释三方面考察协议设计的可靠性(NRC, 2001), 包括评价目标、理论基础、观察视角、数据收集和分析方法。2) 要素层。基于课堂观察的 LICC 范式(崔允漭, 2012)和 STEM 课堂的技术特征, 选取教师教学、学生学习、课堂文化、技术应用四个维度, 对协议的观察项目进行内

表 1 STEM 课堂观察协议选取信息

| 类型 | 协议名称 | 学段 | 选取原因 | 关键引用 |
|-------|--|----------|--|--------------------------|
| 整体性协议 | 改革后的课堂教学观察方案(Reformed Teaching Observation Protocol, RTOP) | K-12, UG | 美国国家科学基金会资助的亚利桑那州卓越教师培训项目(ACEPT)成效评估工具、新协议构建常用的理论模型 | Sawada et al. (2002) |
| | 俄勒冈州教师观察协议(Oregon Teacher Observation Protocol, OTOP) | UG | 美国国家科学基金会资助的俄勒冈州卓越教师培训项目(OCEPT)成效评估工具 | Wainwright et al. (2003) |
| | UTeach 教学观察方案(UTeach Observation Protocol, UTOP) | K-12, UG | 美国国家科学基金会资助的德克萨斯州立大学 UTeach 教师中心培训项目成效评估工具、美国盖茨基金会有效教学评估项目工具 | Walkington et al. (2011) |
| 分段式协议 | 教学维度观察协议(Teaching Dimensions Observational Protocol, TDOP) | UG | 美国国家科学基金会工程和科学教育研究与评估项目(REESE)资助的 STEM 高等教育改革文化、认知和评估研究(CCHER)关键成果 | Hora et al. (2013) |
| | 本科 STEM 课堂观察协议(Classroom Observation Protocol for Undergraduate STEM, COPUS) | UG | 美国国家科学基金会和哥伦比亚大学卡尔·威曼(Carl Wieman)科学教育计划资助的课堂观察项目、Web of Science 高被引工具 | Smith et al. (2014) |
| | 评估主动学习的实践观察准则(Practical Observation Rubric To Assess Active Learning, PORTAAL) | UG | 美国国家科学基金会资助的华盛顿大学第二代主动学习研究(循证教学研究)成果 | Eddy et al. (2015) |

表 2 六套协议分析框架

| 分析层次 | 分析维度 | 分析内容 |
|------|------|------------------------------------|
| 构建层 | 认知 | 协议的评价目标与参考的理论基础 |
| | 观察 | 协议的观察视角和过程性数据的收集 |
| | 解释 | 协议的信效度检验与数据分析方法 |
| 要素层 | 教师教学 | 与教学策略、教学内容、教学实施、教学评价等教师教学相关的具体观察项目 |
| | 学生学习 | 与学生学习投入、认知发展相关的具体观察项目 |
| | 课堂文化 | 与课堂互动、师生关系相关的具体观察项目 |
| | 技术应用 | 与教师使用技术和学生使用技术相关的具体观察项目 |
| 应用层 | 观察主体 | 协议的观察对象 |
| | 评价情境 | 协议的适用性与局限性分析 |

容分析,挖掘其特征要素。3)应用层。围绕观察主体、评价情境两个维度,明确评价过程关注谁以及何时评价。

三、协议比较

(一)构建层:基于评价三角的协议可靠性分析

评价三角框架为验证工具的科学性和有效性提供了理论模型,即任何一种评估工具的设计都必须考虑三个基本维度(NRC, 2001)(见表 3): 1) 认知。这是实施评价的理论基础,用于确定观察任务需要测量的目标; 2) 观察,包括与认知相匹配的观察任务描述,收集与认知理论有关的表现性证据; 3) 解释,指从表现性证据得出合理推论的分析工具和方法,以揭示观察结果如何构成评价结论。在非正式课堂评价中,解释一般由教师依据自身经验,以直观、定性的方式进行;在大规模评估项目中,解释是一种统计模型。在具体评价实践中,认知、观察和解释可以对应理解为评价目标、收集证据和解释证据(赵士果, 2013)。

1. 认知维度

认知维度体现了协议构建的理论基础与评价目标。从教学理念看,三种整体性协议(RTOP、OTOP、UTOP)主张基于建构主义和合作探究的方法进行教学改革,创设以学生为中心、以活动为基础的新型课堂(MacIsaac et al., 2002)。其中,RTOP旨在衡量教学改革的有效性,以探究为导向,考察教师探究式教学方法的运用及学生参与探究的过程;UTOP服务于幼儿园至大学本科阶段的教师培训项目,特别是在 K-12 阶段得到广泛应用,以增值性评价作为选拔优秀教师的标准,重点评估教师课

堂支持和促进学生在学习机会的程度(Walkington et al., 2011); OTOP 用于确定本科课堂使用哪些基于改革的教学要素,并记录改革后的专业发展对学生行为和职前教师教学实践的影响(Wainwright et al., 2003)。然而,整体性协议难以表征课堂的动态变化,只能捕捉相对粗略的教学措施,无法描述具体的教学过程,且涉及大量的教学质量判断,对教师的评分结果可能因观察者的经验不同而产生差异(Hora, 2015)。

三种分段式协议(TDOP、COPUS、PORTAAL)与主动学习导向的本科 STEM 教育改革有关,从教学行为角度揭示课堂的细粒度特征。其中,TDOP 以实践系统理论与分布式认知理论为基础,将观察重点从教师转移至对课堂的全面描述,通过划分课堂实践类别及对应的行为代码衡量教学特征,分析行为主体、技术工具与学习任务间的相互作用(Hora et al., 2013)。COPUS 将互动式教学的实施程度作为衡量主动学习的标准,用于了解师生如何使用课堂时间、确定教师的专业发展需求、验证教学实践调查中教师自我报告的准确性(Smith et al., 2013)。COPUS 采用 TDOP 类似的编码方式,只统计行为频次而无需评级,进而生成可视化的观察结果。PORTAAL 关注教师如何构建课堂经验,帮助教师理解和开展主动学习,基于教育研究的证据确定 21 个在课堂中易于实施且有助于提高学生成绩和逻辑发展能力的主动学习要素(Eddy et al., 2015)。

2. 观察维度

观察维度揭示了协议的观察视角和过程性证据的收集方式。在整体式协议中,RTOP 包含课程设计与实施、教学内容、课堂文化三份分量表共

表 3 基于评价三角的协议可靠性分析

| 工具名称 | 评价目标 | 收集证据 | | | 解释证据 | | | |
|---------|--|--|--|--|------|------|------|--------|
| | | 观察项目 | 数据类型 | 数据收集 | 内容效度 | 标准效度 | 结构效度 | 评价者间信度 |
| RTOP | 评估改革后教学的实施程度, 关注基于探究的教学措施 | 三个维度 25 个观察项目: 课程设计与实施、教学内容(陈述性知识、程序性知识)、课堂文化(师生互动、生生互动) | 五点李克特量表记分(0-4) | 课堂背景和活动记录、观察项目的得分与评分依据 | √ | √ | √ | √ |
| OTOP | 记录改革后本科教学对学生专业发展的影响, 关注改革导向下教学要素的实施频率 | 10 个观察项目: 思维习惯、元认知、学生讨论与协作、富有挑战的想法、先入为主的观念和误解、概念思维、发散思维、跨学科联系、学科知识和概念的多重表述 | 五点李克特量表频次统计(N/O, 1-4), N/O 表示未观察到 | 课堂背景信息、观察项目的得分与评分依据、课堂教学过程的详细记录、访谈工具 O-TIP | √ | | | √ |
| UTOP | 评估 STEM 教师培训与教师专业发展的效果, 关注促进学生学习机会的教学措施 | 四个维度共 26 个观察项目: 课程环境、课程结构、课堂实施和教学内容 | 五点李克特量表记分(1-5), 以 NA 或 NK 表示未观察到和无法判断的项目 | 课堂背景信息、课程概述、各观察项目的得分与评分依据、各维度的整体性评价、观察后的教师访谈 | √ | √ | √ | √ |
| TDOP | 从多维角度描述课堂实践的动态变化, 关注行为主体、教学技术与学习任务间的复杂互动 | 三个基本维度: 教学方法(以教师为中心的方法、以学生为中心的方法)、师生对话(教师主导的对话、学生主导的对话)和教学技术; 三个可选维度: 认知需求、学生参与和教学行为; 共 40 个观察项目 | 2 分钟为间隔, 记录不同代码出现的频次 | 课堂背景信息、标记每 2 分钟出现的所有行为代码、记录重要的课堂活动 | √ | | | √ |
| COPUS | 确定课堂互动式教学的程度, 关注教师与学生如何度过课堂时间 | 25 个基本观察项目: 学生行为(13 个代码), 教师行为(12 个代码)、可选维度: 学生参与度 | 2 分钟为时间间隔, 记录不同代码出现的频次 | 标记每 2 分钟内出现的所有行为代码 | √ | | | √ |
| PORTAAL | 描述课堂主动学习是如何实施的, 关注主动学习最佳实践要素的实施频率 | 四个维度 21 个实践要素: 练习、认知参与、责任制和减少焦虑 | 记录不同代码的持续时间和频次 | 基于包含 41 个项目的观察量表详细记录课堂活动 | √ | √ | | √ |

25 个项目。它基于详细的课堂背景和活动记录, 以五点李克特量表 0(没有发生)-4(非常符合)对观察项目评分, 并注明评分依据供讨论分析。OTOP 归纳了十个改革导向的要素指标, 为每个指标提供了描述性陈述和观察重点说明, 从教师和学生的行为角度描述课堂现象(“概念的多重表述”除外, 该项仅涉及教师行为)。该协议以李克特量表(N/O, 1-4)形式表示不同要素出现的频率, 其数值是分类型而非分值型数据, N/O 代表“未观察到”, 4 代表“经常出现”, 同时需要记录详细的课堂背景信息(课堂类型、课程主题、班级规模)、课堂观察笔记和教师访谈内容, 用于解读 OTOP 结果。UTOP 包括课堂环境、课程结构、课堂实施和教学内容四个维度 26 个项目, 与 RTOP 同为五点李克特量表(1-5)记分式工具。若课堂上某个观察

项目未出现或无法判断, 可以记录为 NA(不适用, Not Applicable)或者 NK(不确定, Not Known), 同时教师还需对每个维度开展五点李克特量表式的整体性评价, 与项目评分形成交叉验证。在分段式协议中, TDOP 涵盖六大实践维度共 40 种行为, 包括三个基本维度(教学方法、师生对话、教学技术)和三个可选维度(教学策略、认知参与、学生参与度), 以 2 分钟为观察间隔, 标记单位时间内出现的所有行为代码; COPUS 将编码减少到 25 个, 仅划分“学生行为”和“教师行为”两个必选维度, 采用与 TDOP 相似的观察间隔与数据记录方式。PORTAAL 将主动学习的 21 个最佳实践要素归纳为练习、逻辑发展、责任制、减少焦虑四个维度, 将课堂分为导入、学生参与、教学环节、汇报环节四个阶段, 编制了包含 41 个项目的课堂观察量表,

详细记录课堂活动发生的时间、频率和事件的存在与否, 依据将观察结果转换为实践要素衡量标准的文档, 得出 21 个要素的实施频率。

上述协议的观察维度主要涉及教学策略、教学内容、教学实施、教学评价、学习投入、认知发展、课堂互动、教学技术八项。首先, 所有协议都关注教学实施、教学评价和课堂互动。教学实施检查教师在课堂实际使用的教学措施, 如提问的策略、开展小组活动的方法、根据学情进行的教学进程调整、讲授学习内容的方式。教学评价考察教师开展学习评估的形式, 包括随堂检测、教师提问与反馈、学生反思与自我评价、学生互评与解释答案等形成性评价手段。课堂互动是课堂文化的重要组成部分, 旨在衡量不同教学主体间的对话与交互情况, 如师生、生生互动的数量和类型。

其次, 除 COPUS 外的协议都关注教学策略与认知发展。教学策略评价教师的教学方法选择和组织能力。例如, RTOP、OTOP、UTOP 观察教师是否考虑学生的原有认知结构设计课堂活动, 帮助学生建立新旧知识的联结; 能否以合作探究的方法创设以学生为中心的课堂结构, 促进学生的沟通交流和思维发展。TDOP 将教学方法分为以教师为中心(如讲授、演示、展示问题解决过程等)和以学生为中心(如小组讨论、自主学习、学生展示等)两类。PORTAAL 通过观察教师提供练习的机会、发展学生高阶思维的方式、鼓励学生参与课堂活动的方法, 评价教师采用的主动学习策略。COPUS 侧重于评估互动式教学的程度, 重点考察课堂的师生行为, 略去对具体行为背后隐含的教学策略的记录。尽管认知发展是一项需要高度推理的内敛指标, 但上述协议仍在外部观察上做了努力。OTOP 以要素提炼的方式将部分观察项目对应到元认知、发散思维、概念思维等认知层面。基于布鲁姆认知领域的教学目标分类, TDOP 划分了解决问题、创新、联系现实等学生认知需求; PORTAAL 考察了课堂练习与考试难度在认知水平上的一致性, 以及课堂中发展的高阶思维类型。RTOP 和 UTOP 虽未将认知发展的观察项目体系化, 但把促进学生思维发展的要求融入教学策略与活动设计、教学内容选择、课堂互动实施等项目的评价。COPUS 删除了这一维度, 因为研发团队发现难以保证精简

协议的同时能可靠地评估学生的认知发展(Lemons et al., 2013)。

除 UTOP 外的五种协议均包含学习投入, 评价学生课堂的参与程度。OTOP 描述了与教学改革要素对应的学生行为, 评估学生是否积极参与有价值的学习活动, 但这种评估方式对观察者要求很高, 因为师生行为被包含在同一协议中, 且需要将双方行为概念化在同一个精细的层次上测量, 难以精准评定(Hora et al., 2013)。为此, TDOP 采用相对粗略的估算方式, 通过统计观察员附近有多少学生在积极做笔记或专注于课堂, 来反映群体的注意力水平和参与程度。这一方法也被 COPUS 沿用。PORTAAL 以“责任感”和“减少焦虑”考察教师鼓励学生参与课堂活动的策略, 关注教师促进学生参与上的努力, 而非学生的实际参与行为。RTOP 以程序性知识评价学生参与探究性活动的情况, UTOP 聚焦教师行为, 并未设立衡量学生参与度的详细项目。

再次, 教学内容维度在 RTOP、OTOP、UTOP 三种整体性协议中均有体现。一方面, 教学内容本身具有很强的连续性, 难以在分段式工具中被特定的行为编码概括; 另一方面, STEM 课堂的教学内容有别于传统课堂, 其认知发展性、概念抽象性、真实情境性都需要单独评价, 例如是否选取了学科核心概念, 采取了何种概念组织与表达的形式, 是否有助于学生理解, 能否与其他学科或现实世界相联系, 以帮助学生进行知识迁移和创新应用。

最后, 教学技术维度在 RTOP、UTOP、TDOP 三种协议中均有提及, 其中 RTOP 与 UTOP 只是将技术的使用简化为对“教师是否使用了合适的教学资源”或者“学生能否使用多种方式和工具表达观点”的判断, 并未讨论技术赋能 STEM 教学的具体过程; TDOP 重点考虑了课堂中信息技术的类型, 将教学技术作为一项单独维度提出, 通过分析其与教学方法、认知参与维度中行为的共现关系, 评价技术的使用特征。

3. 解释维度

解释维度描述了证据分析和解读的过程。证据分析和解读需要考虑数据收集的可靠性和有效性, 同时需要采用合适的统计学模型挖掘数据背后的教育学特征, 将过程性数据意义化和可视化。上述协议的信效度检验与数据分析方法见表 4。

表 4 信效度检验与数据分析方法

| 工具名称 | 信度检验 | | 效度检验 | | | 数据分析 |
|---------|--|---|-------------------------------|---------------------------------|---|---|
| | 评价者信度 | 内部一致性 | 内容效度 | 标准效度 | 结构效度 | |
| RTOP | 最佳拟合线性回归: 整体协议可靠性 0.954; 分量表可靠性: 教学设计与实施 0.915、陈述性知识 0.670、程序性知识 0.946、交流互动 0.907、师生关系 0.872 | 整体协议 Cronbach's α 值为 0.97; 分量表 Cronbach's α 值位于 0.8-0.93 之间 | 基于数学课程教学与评估标准、国家科学教育标准、科学素养标准 | 协议分数与学生成绩提高的相关性为 0.88 | 探索性因子分析: 三个主要因子的累计方差贡献率为 71.92%, 命名为探究取向、陈述性知识和师生关系 | 0-30 分: 传统型教师中心课堂; 31-45 分: 过渡型教师指导式课堂; 46-60 分: 过渡型部分考虑学生的课堂; 61-100 分: 改革型以学生为中心的课堂 |
| OTOP | 项目评分相同和评分相差 1 分的一致性百分比: 元认知 57%、前概念 71%, 其余项目为 100% | 未检验 | 专家咨询 | \ | 未检验 | 基于课堂类型分类讨论 OTOP 项目频次统计结果 |
| UTOP | Kappa 检验: 单个指标的加权 kappa 为 0.41, 综合评分为 0.63 | 分量表 Cronbach's α 值位于 0.905-0.962 之间 | 专家咨询 | UTOP 总分与教师增值的相关性为 0.34 | 探索性因子分析: 四个主要因子为促进表层参与、促进深度理解、跨领域融合、学科知识 | 计算每个观察项目、每个维度、UTOP 总分的平均值与标准差; 项目得分的差异检验 |
| TDOP | Kappa 检验: 五个维度的 kappa 系数位于 0.7-0.9 之间 | \ | 专家咨询 | \ | \ | 利用社会网络分析生成代码之间的共现网络图; 结合 DOLA 框架识别促进“主动学习”的教学措施 |
| COPUS | Jaccard 系数: 89% 代码的 Jaccard 系数在 0.9 以上; Kappa 检验: 平均 kappa 系数位于 0.79-0.87 之间 | \ | 专家咨询 | \ | \ | 生成不同行为时间占比的饼状图; 以聚类分析划分三种教学风格: 灌输式、互动式讲授、以学生为中心 |
| PORTAAL | ICC 组内相关系数为 0.8 | \ | 基于教育研究的证据 | 结构方程模型: RORTAAL 实践要素与学生考试成绩的相关性 | \ | 计算 21 个实践要素的频率或持续时间以及每个实践要素的中位数和四分位数 |

协议的可靠性主要通过考察评价者间的一致性来衡量, 常用 kappa 系数检验定类数据的一致性。UTOP 使用加权 kappa 系数处理量表的等级式定类数据。TDOP 与 COPUS 使用简单 kappa 系数处理行为编码的绝对定类数据, COPUS 采用 Jaccard 系数验证单个行为编码间的一致性。ICC 组内相关系数适用于定量数据间的一致性验证, PORTAAL 借此处理不同评价者统计的课堂实践百分比数据。RTOP 采用回归分析对 16 组观察者的评估数据进行模型拟合, 验证整体协议与各分量表的可靠性。同时, RTOP 和 UTOP 以 Cronbach's α 系数检验量表的内部一致性。OTOP 同为量表式工具, 仅以观察者各指标评分相同或相差 1 分的百分比检验评价者间一致性, 没有基于标准化的统计模型, 也没有对量表进行信度分析, 其可靠性还需检验。

协议的有效性分析包括内容效度、标准效度

和结构效度。1) 内容效度指观察项目与所测量概念或领域的吻合程度, 主要取证于各类标准化文件 (RTOP)、专家意见征询 (OTOP、UTOP、TDOP、COPUS) 或实证研究成果 (PORTAAL)。2) 标准效度指评估结果与外部标准的关联程度, 通常用二者间的相关系数衡量。RTOP 考察了协议得分预测学生成绩提高的能力, 得到相关系数为 0.88, 表明具有较好的预测效度。UTOP 建立了协议总分与教师增值 (与教师相关的成绩增益) 的相关性, 获得教师教学措施与上一年学生成绩增益之间相关性为 $r=0.34$ ^①; 预测同一教师所教另一个班级的成绩的相关性为 $r=0.26$, 均高于有效教学评估项目采用的其他工具 (Kane et al., 2012)。这表明, UTOP 有助于识别对学生成绩提升有明显帮助的优秀教师。由于不能判断每个要素对学生学业成就的相对重要性, PORTAAL 的结果并非单一的

数, 而是教师课堂使用每个要素的平均频率或时间。研发者使用结构方程模型路径分析方法检验 PORTAAL 实践要素与学生考试成绩间的相关性, 发现小组活动、随机点名、解释备选答案、小组合作以及回答问题的总时长与考试成绩呈正相关; 在考核高阶认知能力的试题中, 学生解释答案、独立思考、教师积极反馈等要素也与其考试成绩正相关(Moon et al., 2021)。3) 结构效度表示观察项目解释预期测量的某种结构或特质的程度, 用于检验理论构建的合理性及其转换为测量项目的可操作性。RTOP 与 UTOP 运用探索性因子分析检验协议的结构效度, 但均未开展验证性因子分析。TDOP 和 COPUS 旨在客观描述教学过程, 并非为评估教学质量提供外部标准, 其行为代码也不是为了测量某种潜在构念, 因此其有效性检验没有考虑协议的标准效度和结构效度(Hora, 2015)。

在数据分析上, RTOP 和 UTOP 是量表式计分工具, 需要记录每个项目的分数与协议的整体得分, 然后计算平均值、标准差。RTOP 基于大规模评估实践得出不同学段和学科的常模, 为衡量教学改革程度提供基准; 根据总分, 划分出四种课堂类型: 传统型教师中心式(0-30)、过渡型教师指导式(31-45)、过渡型局部改革式(46-60); 改革型学生主体式(61-100)。UTOP 被证明具有良好的区分度。它借助 T 检验能够分析不同水平教师在具体观察项目上的表现差异。OTOP 的数值是分类型而非分值型数据, 用于讨论改革型教学策略的使用频率, 使用者需要结合详细的课堂背景信息(课堂类型、课程主题、班级规模)和教师访谈内容, 综合描述不同类型课堂(如讲授、实验、演示等)的教学

特征。TDOP 利用社会网络分析生成不同行为之间的共现网络图, 挖掘教学方法、认知参与和教学技术间的相互作用, 分析教学实践系统的核心特征; 同时, 借助差异化公开学习活动(Differentiated Overt Learning Activities, DOLA)框架^②, 捕捉课堂的主动学习行为(Hora, 2015)。COPUS 统计不同行为的持续时间, 以饼状图直观呈现观察结果。有研究者使用聚类分析处理 COPUS 数据, 发现存在三种不同的教学风格: 灌输式、互动式、以学生为中心式(Stains et al., 2018)。PORTAAL 通过计算 21 个观察要素出现的平均频率或持续时间及其中心位数和四分位数, 为教师改善教学实践提供标准。

(二) 要素层: STEM 课堂观察项目的内容分析

要素层旨在挖掘 STEM 课堂观察的特征要素, 把握 STEM 教学实施与评估的重点。由构建层观察维度可知, 上述 STEM 课堂观察协议涵盖教学策略、教学内容、教学实施、教学评价、学习投入、认知发展、课堂互动、教学技术八个方面。崔允漷(2012)构建的面向听评课的课堂观察 LICC(Learning Instruction Curriculum Culture, LICC)范式, 将课堂解构为学生学习、教师教学、课程性质、课堂文化四个方面。在参考国内外典型课堂观察框架的基础上, 考虑到 STEM 课程性质与教师的教学设计和实施紧密相关, 本研究将其并入教师教学维度; 基于教育数字化转型的时代要求和 STEM 课堂的技术增强性(余胜泉等, 2015), 将技术应用作为单独的分析维度。由此, 本研究从教师教学、学生学习、课堂文化、技术应用四个评价维度, 对六套协议的 147 个观察项目进行内容分析, 提炼出 14 项特征要素(见表 5)。

表 5 STEM 课堂观察的特征要素

| 工具名称 | 教师教学 | | | | | 学生学习 | | | | 课堂文化 | | | 技术应用 | |
|---------|------|------|------|------|------|------|------|------|-----|------|-----|-----|--------|--------|
| | 科学探究 | 先验知识 | 沟通合作 | 学科融合 | 实践操作 | 概念思维 | 批判思维 | 创新思维 | 元认知 | 自主性 | 民主性 | 互动性 | 教师使用技术 | 学生使用技术 |
| RTOP | √ | √ | √ | √ | | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| UTOP | √ | √ | √ | √ | | √ | √ | √ | √ | √ | | √ | | |
| OTOP | √ | √ | √ | √ | | √ | √ | √ | √ | √ | √ | √ | √ | |
| TDOP | √ | | √ | √ | √ | √ | √ | √ | √ | √ | | √ | √ | √ |
| COPUS | √ | | √ | | | | | | | √ | | √ | | |
| PORTAAL | √ | √ | √ | | | √ | √ | √ | | √ | √ | √ | | |

1. 教师教学

教师教学维度考察教师的教学策略、教学内容、教学实施和教学评价的表现, 评价教师设计与实施 STEM 课堂教学的能力。具体到观察项目上, 上述协议在此维度均强调科学探究与沟通合作两个要素, 主要体现在教学策略选择与教学实施过程的评价中。科学探究评价教师能否为学生提供模拟真实科学任务的机会, 即能否通过提问、假设、论证、交流、展示等, 引导学生主动参与问题解决过程。沟通合作评估教师如何开展课堂互动和促进学生之间的合作学习, 观察教师能否为学生提供清晰的合作目标和充分的交流机会, 是否允许学生充分表达观点并给予正向反馈与引导, 推动学习共同体的形成。其次, 先验知识与学科融合被四种协议提及。先验知识关注教师的教学策略与活动选择有无考虑学生的已有知识与概念, 调整教学内容使之符合学生的认知发展水平, 帮助其建立新旧知识的联结。学科融合是评价教学内容设计的关键指标, 体现 STEM 教育的学科融合性与真实情境性, 要求教师将教学内容与其他学科及真实世界相关联, 基于真实情境讨论问题解决策略, 促进学生知识的迁移应用与创新发展。最后, 实践操作层面的观察项目不足, 即忽视课堂上原型制作、产品测试、模型迭代等环节的评价。只有 TDOP 提出“教师应要求学生产生创造性的想法与产品”, 但这一观察项目没有落实到学生思维的可视化与产品化过程, 难以体现 STEM 课堂的工程设计与思维元素。

2. 学生学习

学生学习维度包括学习投入和认知发展, 评价学生在 STEM 课堂的外显学习行为和可能的思维发展过程。上述协议对学习投入的评价主要通过程序性知识观察学生知识的获取过程, 例如, 从处理信息、得出结论和评估知识等方面检查学生是否参与了科学探究活动, 或者以百分比形式粗略计算学生的参与程度(如高、中、低)。认知发展与学生的课堂参与行为紧密相关, 旨在评价课堂对学生高阶思维培养的促进作用, 主要包括概念思维、批判思维、创新思维和元认知能力, 这与已有研究的观点高度一致(首新等, 2021; 阙荣辉等, 2023)。概念思维是关联性思维, 是高阶思维发展的基础(朱彩兰等, 2023), 其培养与陈述性知识的传授密

切相关, 需要教师的学科知识水平与内容组织能力, 包括如何从不同视角对同一概念进行多样化解读, 如何组织跨学科主题下的不同概念, 能否使用图表或模型帮助学生厘清知识间的本质联系; 还需观察学生的观点表达与作品制作有无运用多学科知识解决问题, 将所学内容与上位概念联系起来理解。批判思维、创新思维和元认知能力的观察主要关联程序性知识, 评价教学活动设计的合理性和学生参与的有效性, 关注教学活动是否涉及预测、估计、假设、协商等科学思维方式, 或者应用、分析、综合、评价等高阶认知技能; 评价学生在课堂中能否基于证据严谨地思考、批判性地审视观点、在建构性对话中生成新的想法, 以及能否基于总结和反思活动进行自我调节。

3. 课堂文化

课堂文化维度反映师生共同营造的社会性氛围以及整个教学过程传递的价值观信息。上述协议主要以课堂互动体现这一维度的特征, 包括生生互动和师生互动两方面。前者关注学生之间互动的数量和类型, 以及教师如何促进这种互动, 包括学生参与课堂讨论与小组活动的方式、学生与他人互动的态度(如善于自主表达、能够尊重他人、敢于批判质疑、乐于协商共享); 后者关注师生关系和教师营造课堂文化的理念, 评价教师是否以“听众”或“顾问”的身份参与, 从不可侵犯的权威变为引导学生的交流对象与合作伙伴。整体来看, 所有协议都体现了 STEM 课堂文化的自主性与互动性特征, 评价课堂是否为学生提供独立思考与自主学习的机会, 以及教学过程如何在教师的提问与反馈、学生的思辨与讨论中双向进行。其中, RTOP 和 UTOP 还关注教师是否根据学生的观点和表现调整教学进度、分配课堂时间、有效处理突发情况, 评价教师 STEM 课堂管理的质量, 为生成式、探索性的学习环境提供保障; TDOP 描述了人机互动与师生互动、生生互动共同构成的课堂互动文化。此外, RTOP、OTOP、PORTAAL 还评价 STEM 课堂的民主性特征, 关注减少学习焦虑的教学措施, 例如给予学生主动提问的机会、鼓励学生提出不同想法、允许犯错并肯定错误的价值。

4. 技术应用

虽然上述协议的观察项目并未详细讨论技术

支持 STEM 课堂教与学的方式,但仍可以从技术应用项目中获得启发。技术应用维度主要考虑学生和教师使用技术两个方面,前者评估技术是否作为学生的认知工具,培养学生运用技术解决问题的能力(余胜泉等, 2015),如学生使用多种技术媒介与他人讨论交流;后者评估教师的技术使用行为,判断其是否优化了学生知识建构与能力发展的过程(宋歌等, 2022),如教师能否获取并使用适切的教学资源,重点观察教师如何使用技术提高学生学习效果。同时,使用的技术类型也是这一维度的重要评价指标,如 TDOP 列举了教师可能使用的教学技术,包括电子白板、PPT、答题器、演示设备、个人平板、视频、虚拟仿真系统、网站等,但以上分类仍以传统的演示型技术和交互型技术为主,难以充分反映 STEM 课堂的技术增强特征。

(三)应用层:协议的实践特征、优势与局限

正式实施课堂观察应明确不同工具的观察主体和评价情境。观察主体区分了协议的关注对象,即是以教师为主还是以学生为主;评价情境涉及协议的作用类型、适用范围与局限性。六套协议的应用特征见表 6。

1. 观察主体

从观察主体看,UTOP 与 PORTAAL 是典型的以教师行为为焦点的协议,其观察项目仅考察教师可能采用的有效教学措施,没有直接记录课堂中学生的反应。RTOP 与 TDOP 关注教师行为,同时在个别维度观察学生行为。RTOP 在“程序性知识”

中考察学生参与科学探究的过程,在“生生互动”中评估学生互动的数量和类型;TDOP 在“师生对话”与“学生参与”中描述学生的主动对话与专注度。上述协议仅有 OTOP 与 COPUS 同时以教师和学生为观察焦点,OTOP 提出不同实践要素对应的课堂学生行为表征;COPUS 特别将“学生行为”置于协议首部,以防观察员习惯性地关注教师行为而忽视课堂的真正主体——学生。

2. 适用性

从作用类型看,RTOP、UTOP、PORTAAL 属于评估性工具,可建立常模以评估课堂教学质量,并在一定程度上预测学生成绩;OTOP、TDOP、COPUS 为描述性工具,观察结果作为一种分类数据,旨在客观描述课堂教学实践。不同的评价目标与内容结构直接影响协议的适用范围。RTOP、OTOP、UTOP 能够评估 STEM 课堂的教学改革程度,是中小学教师专业发展的有效工具。其中,RTOP 面向广泛的教学方法评估与反馈,能衡量各类教改措施的有效性,例如引导探究(Blanchard et al., 2010)、协作探究(Cavagnetto et al., 2011),师生对话互动等教学策略(Akkus et al., 2011),以及以学生为中心编制的教学资源(Teasdale et al., 2020)。UTOP 考虑了从探究式教学到讲授式教学的不同教学模式,有助于了解教师教学的优劣势,找出其薄弱环节。OTOP 是职前教师开展教学计划和实践反思的有效工具,能促进其更快适应正式教学工作(Bullock et al., 2012)。TDOP 能描述教学的动态

表 6 六套协议的应用特征

| 协议名称 | 观察主体 | | 评价情境 | | |
|---------|------|----|------|-------------------------------------|--|
| | 教师 | 学生 | 类型 | 适用情境 | 局限性 |
| RTOP | √ | 部分 | 评估性 | 评估改革后教学措施的有效性;评估课堂教学质量;教师专业发展;同行评议 | 无法直接测量学生学习结果;未提供量表每个级别分数的具体描述;培训时间长 |
| OTOP | √ | √ | 描述性 | 同行评议;职前教师专业发展 | 依赖于现场笔记;对评价者观察经验要求高;没有考虑学生的消极反应 |
| UTOP | √ | | 评估性 | 评估课堂教学质量;教师专业发展;优秀教师选拔 | 难以捕捉学科思维发展证据;广泛的观察维度可能导致评价失焦;培训时间长;缺乏对学生行为的观察 |
| TDOP | √ | 部分 | 描述性 | 记录课堂教学实践类型;捕捉课堂动态变化;评估教学干预效果;教师专业发展 | 行为编码存在冗余;培训时间长;无法描述非代码库行为;难以充分捕捉学生行为 |
| COPUS | √ | √ | 描述性 | 描述课堂的教学实践;提供师生课堂时间分配的反馈;教师专业发展 | 数据计算公式有待修正;忽略课堂的动态变化;教学风格分类不够充分;没有考虑学生的消极反应 |
| PORTAAL | √ | | 评估性 | 评估课堂开展主动学习的质量;教师专业发展 | 没有描述学生的特征;只能捕捉既定的课堂实践;不适用于实验课程与小型研讨课;难以在课堂实时使用 |

变化,借助社会网络分析厘清不同实践维度内部与之间的关系,从行为、认知、技术三方面挖掘不同学科与课型的教学特征。COPUS 突出学生的课堂主体地位,可用于分析师生互动、直观了解师生如何使用课堂时间;其代码的精简性便于没有观察经验或科学教育研究背景的中小学 STEM 教师、大学教师与教研人员经短期培训后使用。RORTAAL 将主动学习实践解构为课堂可操作性行为,为教师改革课堂实践提供参考框架,有助于主动学习的理解与实施。

3. 局限性

没有一种工具能够满足所有课堂教学评估的需求。上述协议也存在局限,主要体现在协议设计、评价效果与易用性三方面。在协议设计上,RTOP、UTOP、TDOP、PORTAAL 没有充分关注学生行为;COPUS 只考察了学生行为的类型,不评估学生课堂参与的质量。同时,上述协议以学生的积极表现为评价重点,没有考虑学生的抵抗行为或消极反应 (Shekhar et al., 2015)。在评估效果上,仅有 TDOP 考虑课堂的动态变化和不同实践维度间的关系。RTOP、OTOP、UTOP 是量表式的静态描述,只给出整体的分值,省略了具体教学行为发生的时间;COPUS 与 PORTAAL 虽然同为分段式协议,但其观察间隔的划分多为了统计不同行为课堂出现的频次,没有探索行为间的相互作用。此外,PORTAAL 倾向于描述大型的讲授式课堂,难以反映小型研讨课与实验课的教学特征,因为这类课程的生生互动频次与深度显著增加,还伴随着较多的动手操作环节。

从易用性看,所有协议正式使用前,均需要开展前期培训,以便达到评价者间的一致性。除了 COPUS 与 PORTAAL 的培训时间较短(1.5 h 与 5 h),其余协议都需较高的时间成本。使用手册的不完备也为观察者带来困扰。其中,RTOP 没有为李克特量表的每个级别提供具体描述,不容易解释中间量级的分差。OTOP 需要结合师生两个行为主体对每个维度进行综合评定,在很大程度上依赖详细的现场笔记;协议涉及的元认知、前概念等要素也对观察者的教学经验和理解能力提出了较高要求。UTOP 并未归纳二级维度,试图涵盖众多教学行为的做法可能导致局部评价失焦。TDOP 部分编码

的合理性受到质疑,被认为存在冗余 (Smith et al., 2013),无法捕捉非编码库中的行为,难以描述真实课堂中学生的参与程度和注意力分布。COPUS 的数据处理方法有待完善,其饼状图是某一行为被勾选的频次除以被勾选行为的总数得到的百分比,用以表示课堂某种行为的时间比例,但这种计算方式容易受编码总数的影响 (Lund et al., 2015)。此外,COPUS 教学风格分类难以区分教师形成性评估与反馈的差异,特别是对互动式讲授和以学生为中心的课堂 (McConnell et al., 2021)。PORTAAL 的不同要素对应特定的观察项目,数据统计涉及繁琐的数据转换,需要观察员投入更多的时间与精力。

四、特征要素与评价矩阵

(一) 特征要素

基于对六套典型 STEM 课堂观察协议研发范式、构成要素和实践特征的系统分析,结合 STEM 课堂学科融合性、真实情境性、项目承载性的特点,本文以探究共同体框架 (Community of Inquiry Framework, COI 框架) (Garrison et al., 1999) 和 TPACK 框架为理论依据,将 STEM 课堂解构为教学存在 (Pedagogical presence)、社会存在 (Social presence)、技术存在 (Technological presence) 三个维度,以完善上述协议实际应用的局限 (如对学生学习观测不足、忽视教学技术的作用、课堂实践操作特征不明显等),并基于教师与学生两个观察主体,统摄上述协议在教师教学、学生学习、课堂文化、技术应用四个评价范畴涉及的八个课堂观察维度和观察项目体现的 14 项特征要素,形成 STEM 课堂观察的 L-PST (Learning-Pedagogical Social Technological) 双罗盘模型 (简称 L-PST 模型) (见图 1)。

L-PST 模型以教学存在、社会存在和技术存在三个维度解构 STEM 课堂的教与学行为,同时考虑教师和学生两个观察主体,划分为内外双层罗盘。其中,学生学习是 STEM 课堂的中心,位于模型内环 (观察学生行为表现,称为“学生罗盘”),包括学生的认知发展、社会参与、使用技术;教师教学是影响学生学习结果的关键因素,位于模型外环 (观察教师行为表现,称为“教师罗盘”),包括教师课堂教学、创设课堂文化、使用技术。

教学存在表征教师对教学系统的设计组织及

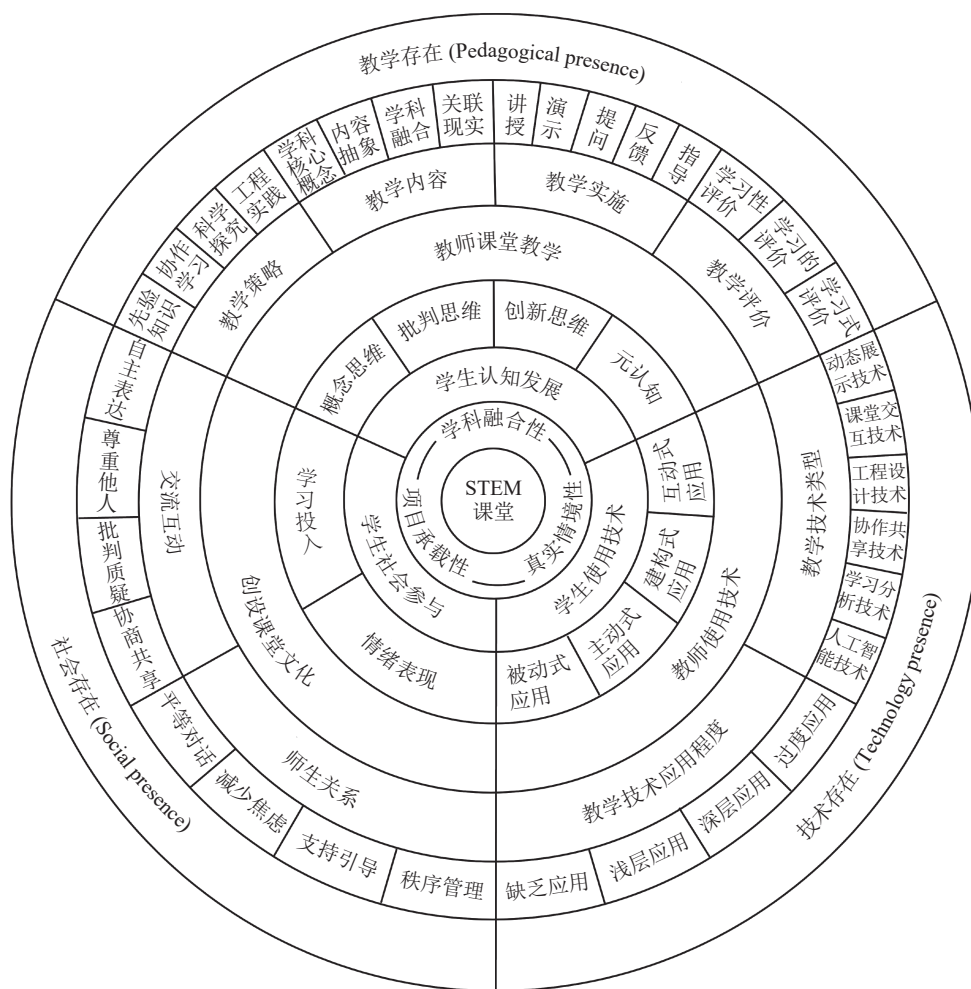


图1 STEM 课堂观察的 L-PST 双罗盘模型

对学生的引导方式,以促成有价值的学习成果,关注学生认知发展和教师课堂教学。学生认知发展指向实践探究的高阶思维培养,评估学生意义主动建构与迁移创造过程。STEM 课堂的思维培养目标主要包括概念思维、批判思维、创新思维和元认知能力。概念思维旨在评价学科大概念与跨学科知识体系建构;批判思维强调能够基于证据意识进行严谨全面地思考;创新思维重视生成新想法、新方案与新产品的能力;元认知指向学习的监控与调节过程。教师课堂教学评价教师设计、组织与实施 STEM 课堂的能力,涵盖教学策略、教学内容、教学实施、教学评价四个子维度。教学策略以先验知识、协作学习、科学探究、工程实践为评价指标,观察教师是否考虑学生先验知识的影响,是否采用合作探究的教学方法发展学生的认知结构,能否以工程实践活动落实学生的创新想法并提升其

动手操作能力。教学内容评价教师如何促进学生对概念的系统性理解,包括学科核心概念与原理的选择、内容表达的抽象化与多样性、将教学内容与其他学科和现实世界关联的有效性。教学实施基于五种教师行为类型(讲授、演示、提问、反馈、指导)(Lund et al., 2015),评价教师组织与实施 STEM 课堂的方式。教学评价观察教师能否有意识地收集学生跨学科实践的表现性证据(学习性评价),根据学习表现进行即时反馈与调整教学,引导学生开展自我评价与学习反思(学习式评价),并基于阶段性学习结果(学习的评价),对学习成效进行合理测评(詹泽慧等, 2023)。

社会存在表征学生的社会参与和教师创设的课堂文化,旨在营造促进探究、质疑、分享的主动学习环境。学生社会参与评价学生的课堂参与程度,内涵包括学习投入(如认知投入和社会投入)和

情绪表现两个子维度。学习投入基于 TDOP 识别主动学习行为的 ICAP 框架 (Chi et al., 2014), 分为被动学习、主动学习、建构学习和互动学习四种状态。被动学习指学生只是被动地从教学材料中接受信息 (如学生被动地听和看、跟随教师讲解抄写笔记、阅读学习资料); 主动学习表示学生积极参与教学, 主动使用学习材料 (如旋转物体进行观察、自主标注重要句子); 建构学习强调学生能超越学习材料生成新的想法与产品 (如提出问题与假设、自我评价、构建模型与调试); 互动学习指学生在建构性对话中生成新的想法 (如整合同伴的观点、与同伴展开辩论)。情绪表现是影响学生社会性参与的决定性特征, 可分为消极情绪和积极情绪。上述协议只评价学生的积极参与行为。然而, STEM 课堂上学生往往承担很多面向真实情境的复杂任务, 会引发对自身能力的焦虑。因此, 还需关注学生对 STEM 教学活动和主动学习策略的抵抗反应, 包括学生是否呈现消极的面部表情 (被动的、非语言的抵抗)、敷衍地完成任务 (部分抵抗)、公开反对参与任务 (公开抵抗)。创设课堂文化维度考察交流互动和师生关系。前者考量教师如何为学生营造有助于意义建构的课堂氛围以及引导学生协作的方法, 观察学生课堂互动的方式和态度, 评价学生是否有自主探索和自主表达的机会和意愿、尊重他人表达的权力、能否基于证据批判性地看待同伴和教师的观点、在协商共享中实现群体智慧的共创。后者考量教师能否作为知识获取的协助者、思维发展的引导者、课堂制度的管理者, 为课堂教学创设自主、民主、互动的氛围。

技术存在表征师生在 STEM 课堂的技术应用类型和程度, 关注技术赋能教与学的过程。学生使用技术关注技术促进学生主动参与和深度学习的方式, 与学习投入的四种状态对应, 如学生跟随教师指令在平台查看学习材料 (被动型)、使用搜索引擎收集信息 (主动型)、设计虚拟仿真模型 (建构型)、协作绘制思维导图 (互动型)。“教师使用技术”考察课堂的技术资源类型和教师技术应用水平。教学技术类型评估教师是否理解 STEM 课堂的技术特征, 能否选用合适的教学工具。基于美国联邦教育部报告的九种 STEM 教育技术赋能方式 (Office of Educational Technology, 2019), 本研究根

据表现形式和作用类型将 STEM 课堂常见的技术资源分为动态展示技术 (虚拟仿真系统、数据可视化模型等)、课堂交互技术 (Kahoot、雨课堂、UMU 平台等)、工程设计技术 (图形化编程、开源硬件、激光切割、3D 打印等)、协作共享技术 (协作思维导图、项目式数字协同平台)、学习分析技术 (学习管理系统、传感器与测量设备等)、人工智能技术 (个性化资源推送、ChatGPT 类技术等) 六种。教师技术应用水平衡量教师的技术应用适切程度, 涉及缺乏应用 (教师很少/不愿意在 STEM 课堂使用技术)、浅层应用 (以教师为中心使用技术, 如基于技术简单呈现教学内容)、深度应用 (以学生为中心使用技术, 如基于技术促进学生的项目式学习和个性化学习)、过度应用 (技术使用的目的含糊、效果不佳, 如使用超出学生认知水平的复杂技术, 阻碍正常的教学进程) 四个层级 (Jen et al., 2016)。

总体而言, STEM 课堂观察的特征要素须体现 STEM 教育的学科融合性、真实情境性与项目承载性。在学科融合性上, 教学内容维度强调对学科核心概念选取、内容本质抽象、跨学科迁移等学科整合属性的评价; 认知发展维度强调观察学生概念思维的发展。真实情境性由教学设计维度的先验知识要素和教学内容维度的联系现实要素表征, 评价教师能否考虑学生原有认知结构, 围绕社会生活创设问题情境, 从而强化学生概念的迁移应用与情境性生成, 推动创新思维的发展。项目承载性体现在教学策略维度下的协作学习、科学探究、工程实践, 考察教师开展项目式教学的设计能力, 以学生社会参与和教师创设的课堂文化维度评价项目式学习的实施程度。整个教学过程中, 观察者还需要考察教师是否实施了多元的教学评价措施, 并评估技术对教与学过程的影响。

(二) 三维矩阵

本文结合典型 STEM 课堂观察协议的应用特征, 以评价三角理论为框架, 构建了 STEM 课堂观察评价的三维矩阵 (见图 2)。评价三角理论的三个要素彼此关联, 形成 STEM 课堂观察实施的三个方面:

一是认知维度与观察维度构成的课堂观察设计层。观察者依据教育教学理论, 确定评价目标与内容, 从 STEM 课堂观察特征要素中选取对应视角,

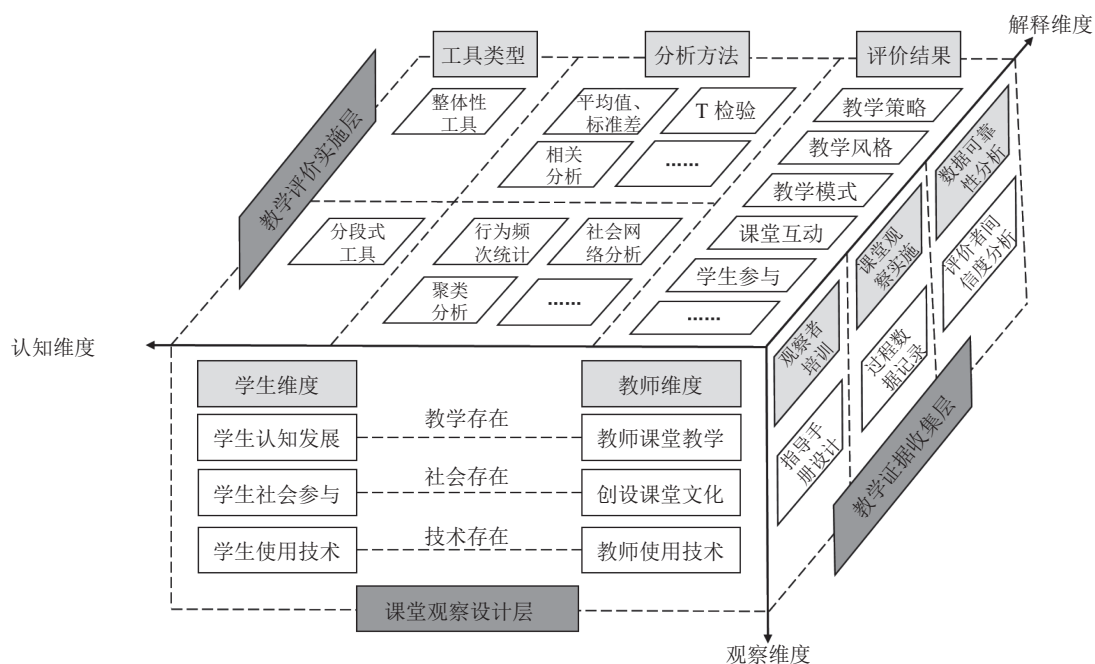


图2 STEM 课堂观察评价的三维矩阵

使用已有协议或自行编制整体性或分段式工具。例如,考察教师行为对学生参与度的影响,可基于教师教学的五种教学实施行为,结合其他子维度的特征要素进行扩展,如教师讲授行为可以包括陈述板书(L1)、联系旧知(先验知识, L2),教师提问可包含提出封闭性问题(Q1)、联系日常生活提出开放性问题(关联现实, Q2);学生学习投入可基于被动、主动、建构、互动四个层级划分。

二是观察维度和解释维度构成的教学证据收集层。观察者基于观察内容与数据分析要求,详细记录课堂教学过程,并将其量化为可处理的教学数据。在此过程中,观察者需先接受培训,基于观察指标的评价依据,包括设计目的、评价重点、行为表征、程度说明等,帮助教师和教研员理解 STEM 课堂评价的重点;其次,使用评价工具开展课堂观察与证据收集,除核心行为数据,辅助性的质性材料,如课堂过程的描述性笔记、教师访谈、课堂背景等信息,也可作为数据解读的参考;最后,分析评价者间的信度,证明数据的可靠性与可用性。

三是认知维度与解释维度构成的教学评价实施层。观察者借助多元的解释方法将教学数据转化为可理解的评价结果,以评价目标涉及的认知理论为基础解读教学证据,得出可靠的推理与结论。例如,量表式计分工具可用于计算不同项目与维度

得分的平均值、标准差,分析教师不同指标的表现水平,提出改进建议;T 检验可用于比较教师群体在哪些教学维度或项目上存在显著差异,证明干预的有效性或总结教师专业发展的经验;相关分析或回归分析可用于建立评价结果与教师增值、学生成绩等方面的联系,探讨影响课堂教学质量的关键因素,验证评价工具的可预测性;分段式工具可用于描述性统计不同行为编码出现的频次,从整体上评价课堂教学特征;社会网络分析或滞后序列分析可用于考察不同行为间的共现关系或因果关系,可视化隐性的关键行为序列;行为数据聚类可用于挖掘可能的教学模式与教学风格。此外,观察者还可结合不同类型的工具,如利用 RTOP 评估教学质量,用 COPUS 分析互动特征等,探讨优质课堂与传统课堂的教学行为差异(Lund et al., 2015)。

五、结语

证据导向的评价能够兼顾过程与结果,实现以评定教、以评促学的双重目的(詹泽慧等, 2022)。本文立足 STEM 教育的循证评价理念,以课堂观察为方法,从理论构建、项目分析、实践应用三方面系统梳理和分析国际典型的 STEM 课堂观察协议,明确 STEM 课堂观察维度的构成要素;结合实证研究与大规模评估项目,分析不同工具在教学策略评

估、教学实践识别、课堂互动分析、教师能力选拔等方面的适用性, 以及在学生参与行为与认知发展观测、教学技术使用、实践操作特征挖掘等方面的局限。在此基础上, 本文以探究共同体理论和 TPACK 框架为依据, 将 STEM 课堂解构为教学存在、社会存在、技术存在三个维度, 围绕教师和学生两个观察主体, 形成 STEM 课堂观察的 L-PST 双罗盘模型; 以学生学习为中心, 以教师教学、课堂文化和教学技术为关键, 从六个一级维度八个二级维度 44 个观察要点解构 STEM 课堂观察的特征要素; 基于评价三角理论, 从评价设计、证据收集、评价实施三方面构建 STEM 课堂观察评价的三维矩阵, 旨在为当前 STEM 课堂教学评价提供系统化方法。STEM 课堂教学评价是一个复杂、精细的过程, 尤其是教与学的动态评估依赖于大量的教学证据(余胜泉等, 2019)。因此, 尽管课堂观察在提供直接证据上具有优势, 但仍需整合其他评价方法挖掘外显行为背后隐含的认知与情感因素。

[注释]

① 学生成绩增益是评价教师对学生学习实际贡献的不完美标准, 因为这种增益可能与与教师实践无关的其他因素引起。然而, 由于这种测量误差一直存在, 年与年之间的相关性被削弱了。为实现对基本技能的不完美测量而非技能本身, 此处采用的测量方式的任何一个测度与基础得分之间的相关性都是两个误差测度之间相关性平方根。当报告一年度到另一年度的学生成绩提高的相关性为 0.20 ~ 0.50 时, 这就等于说任何一年度观察到的平均学生成绩提高与教师基础增值之间的相关性在 0.45~0.7 之间。如果观察到的学生成绩增长是几年的平均数, 那么这种相关性会更强。

② 差异化公开学习活动 (Differentiated Overt Learning Activities, DOLA) 框架是一种基于可观察的学生课堂行为构建的主动学习模式的分类方法, 其核心特征之一是 ICAP 理论: 将学生的课堂参与分为三种递进类型: 主动型(学生明显参与激活自己知识的活动)、建设型(学生明显参与生成自己知识的活动)和互动型(两个或多个学生明显参与发展知识的活动)。

[参考文献]

[1] Akkus R., & Hand B.(2011). Examining teachers' struggles as they attempt to implement dialogical interaction as part of promoting mathematical reasoning within their classrooms[J]. *International Journal of Science and Mathematics Education*, 9(4): 975-998.

[2] American association for the advancement of science.(2012). Describing and measuring undergraduate STEM teaching practices[J].

Retrieved April, 10: 2018.

[3] Blanchard M. R., Southerland S. A., Osborne J. W., Sampson V. D., Annetta L. A., & Granger E. M.(2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction: Guided Inquiry and Verification Laboratory Instruction[J]. *Science Education*, 94(4) : 577-616.

[4] Bybee R. W. (2013). The case for stem education: Challenges and opportunities[M]. NSTA Press: 82-87.

[5] Bullock S. M., & Russell T. (2012). Self-Studies of science teacher education practices: Vol. 12[M]. Dordrecht: Springer Netherlands: 157-173.

[6] Cavagnetto A. R., Hand B., & Norton-Meier L.(2011). Negotiating the inquiry question: A comparison of whole class and small group strategies in grade five science classrooms[J]. *Research in Science Education*, 41(2): 193-209.

[7] Chi M. T. H., & Wylie R.(2014). The ICAP framework: Linking cognitive engagement to active learning outcomes[J]. *Educational Psychologist*, 49(4): 219-243.

[8] Council N. R. (2001). Knowing what students know: The science and design of educational assessment[M]. National Academies Press: 44-51.

[9] 崔允漷(2012). 论课堂观察 LICC 范式: 一种专业的听评课[J]. *教育研究*, 33 (5): 79-83.

[10] Ebert-May D., Derting T. L., Hodder J., Momsen J. L., Long T. M., & Jardeleza S. E.(2011). What we say is not what we do: Effective evaluation of faculty professional development programs[J]. *Bioscience*, 61(7): 550-558.

[11] Eddy S. L., Converse M., & Wenderoth M. P.(2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes[J]. *Cbe-Life Sciences Education*, 14(2): ar23.

[12] English L. D.(2016). STEM education K-12: Perspectives on integration[J]. *International Journal of STEM Education*, 3(1): 3,s40594-016-0036-1.

[13] Erdogan N., Navruz B., Younes R., & Capraro R. M.(2016). Viewing how stem project-based learning influences students' science achievement through the implementation lens: a latent growth modeling[J]. *Eurasia Journal of Mathematics Science and Technology Education*, 12(8): 2139-2154.

[14] Garrison D. R., Anderson T., & Archer W.(1999). Critical inquiry in a text-based environment: computer conferencing in higher education[J]. *The Internet and Higher Education*, 2(2-3): 87-105.

[15] Hora M. T.(2015). Toward a descriptive science of teaching: how the ttop illuminates the multidimensional nature of active learning in postsecondary classrooms[J]. *Science Education*, 99(5): 783-818.

[16] Hora Matthew Tadashi, & Ferrare J. J.(2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching[J]. *Journal of the Learning Sciences*, 22(2): 212-257.

- [17] Hora Matthew T, Oleson A., & Ferrare J. J. (2013). Teaching dimensions observation protocol (tdop) user's manual, Madison: Wisconsin Center for Education Research, University of Wisconsin Madison. <http://tdop.wceruw.org/Document/TDOP-Users-Guide.pdf>.
- [18] 胡艺龄, 张启迪, 吴怵(2022). 多元评价视域下 STEM 课堂教学特征与互动模式分析 [J]. 中国远程教育, 572 (9): 44-52.
- [19] Jen T. H., Yeh Y. F., Hsu Y. S., Wu H. K., & Chen K. M.(2016). Science teachers' TPACK-Practical: Standard-setting using an evidence-based approach[J]. Computers & Education, 95: 45-62.
- [20] Kane T. J., & Staiger D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research Paper. MET Project.[J]. Bill & Melinda Gates Foundation: 45-48.
- [21] Lemons P. P., & Lemons J. D.(2013). Questions for Assessing higher-order cognitive skills: it's not just bloom's[J]. CBE—Life Sciences Education, 12(1): 47-58.
- [22] Lund T. J., Pilarz M., Velasco J. B., Chakraverty D., Rosploch K., Undersander M., et al. (2015). The Best of Both Worlds: Building on the COPUS and RTOP Observation Protocols to Easily and Reliably Measure Various Levels of Reformed Instructional Practice[J]. Cbe-Life Sciences Education, 14(2).
- [23] 刘晟, 魏锐, 周平艳, 师曼, 王郢, 刘坚等(2016). 21 世纪核心素养教育的课程、教学与评价 [J]. 华东师范大学学报(教育科学版), 34 (3): 38-45+116.
- [24] MacIsaac D., & Falconer K.(2002). Reforming physics instruction via rtoP[J]. The Physics Teacher, 40(8): 479.
- [25] McConnell M., Boyer J., Montplaisir L. M., Arneson J. B., Harding R. L. S., Farlow B., et al.(2021). Interpret with Caution: COPUS Instructional Styles May Not Differ in Terms of Practices That Support Student Learning[J]. CBE—Life Sciences Education, 20(2): ar26.
- [26] Moon S., Jackson M. A., Doherty J. H., & Wenderoth M. P.(2021). Evidence-based teaching practices correlate with increased exam performance in biology[J]. PLOS ONE, 16(11): e0260789.
- [27] Office of educational technology, us department of education. (2019). Innovation spotlights: Nine dimensions for supporting powerful stem learning with technology[EB/OL]. [2019-10-07]. <https://tech.ed.gov/files/2019/10/stem-innovation-spotlights-research-synthesis.pdf>.
- [28] 秦瑾若, 傅钢善(2017). STEM 教育: 基于真实问题情景的跨学科式教育 [J]. 中国电化教育, (363): 67-74.
- [29] 阙荣辉, 周丽芳(2023). STEAM 教育跨学科思维评价研究——基于化学师范生的实证分析 [J]. 化学教育(中英文), 44 (6): 59-66.
- [30] Rosenberg M. B., Hilton M. L., & Dibner K. A. (2018). Indicators for Monitoring Undergraduate STEM Education[M]. Washington, D. C.: National Academies Press: 22.
- [31] Sawada D., Piburn M. D., Judson E., Turley J., Falconer K., Benford R., et al.(2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol[J]. School Science and Mathematics, 102(6): 245-253.
- [32] Shekhar P., Demonbrun M., Borrego M., Finelli C., Prince M., Henderson C., et al.(2015). Development of an observation protocol to study undergraduate engineering student resistance to active learning[J]. International Journal of Engineering Education, 31(2): 597-609.
- [33] Smith M. K., Jones F. H. M., Gilbert S. L., & Wieman C. E.(2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university stem classroom practices[J]. Cbe-Life Sciences Education, 12(4): 618-627.
- [34] Smith M. K., Vinson E. L., Smith J. A., Lewin J. D., & Stetzer M. R.(2014). A campus-wide study of stem courses: New perspectives on teaching practices and perceptions[J]. CBE—Life Sciences Education, 13(4): 624-635.
- [35] Stains M., Harshman J., Barker M. K., Chasteen S. V., Cole R., DeChenne-Peters S. E., et al.(2018). Anatomy of STEM teaching in North American universities[J]. Science, 359(6383): 1468-1470.
- [36] 首新, 黄秀莉, 李健, 胡卫平(2021). 基于 STEM 学习目标的高阶思维评价 [J]. 现代教育技术, 31 (3): 20-27.
- [37] 宋歌, 管玉琪(2022). 面向整合式 STEM 的教师跨学科素养: 结构模型与发展路径 [J]. 现代远程教育研究, 34 (3): 58-66.
- [38] 宋乃庆, 刘燕玉, 高鑫(2022). 我国 STEAM 教育的研究现状与发展路径——基于文本聚类的文献计量分析 [J]. 中国教育科学(中英文), 5 (4): 137-147.
- [39] Teasdale R., Ryker K., Viskupic K., Czajka C. D., & Manduca C.(2020). Transforming education with community-developed teaching materials: Evidence from direct observations of STEM college classrooms[J]. International Journal of Stem Education, 7(1): 56.
- [40] Wainwright C., Flick L., & Morrell P.(2003). The development of instruments for assessment of instructional practices in standards-based teaching[J]. Journal of Mathematics and Science: Collaborative Explorations, 6(1): 21-46.
- [41] Walkington C., Arora P., Ihorn S., Gordon J., Walker M., Abraham L., et al. (2011). Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program (UTeach Technical Report 2011-01). Austin: UTeach Natural Sciences, University of Texas at Austin.
- [42] 许娜, 高巍, 郭庆(2020). 新课改 20 年课堂教学评价研究的逻辑演进 [J]. 教育研究与实验(6): 49-55.
- [43] 余胜泉, 胡翔(2015). STEM 教育理念与跨学科整合模式 [J]. 开放教育研究, 21 (4): 13-22.
- [44] 余胜泉, 吴澜(2019). 证据导向的 STEM 教学模式研究 [J]. 现代远程教育研究, 31 (5): 20-31+84.
- [45] 詹泽慧, 季瑜, 赖雨彤(2023). 新课标导向下跨学科主题学习如何开展: 基本思路与操作模型 [J]. 现代远程教育研究: 1-10.
- [46] 詹泽慧, 姚佳静, 吴倩意, 黄秉刚(2022). 人工智能课程中表现性评价的设计与应用 [J]. 现代教育技术, 32 (5): 32-41.
- [47] 张屹, 赵亚萍, 何玲, 白清玉(2017). 基于 STEM 的跨学科教学设计与实践 [J]. 现代远程教育研究(6): 75-84.
- [48] 朱彩兰, 陈彤, 李艺, 沈书生(2023). 关联思维的内涵与形成路径研究 [J]. 电化教育研究, 44 (5): 29-35+43.
- [49] 赵士果(2013). 促进学习的课堂评价研究 [D]. 上海: 华东师范大学: 96-102.

(编辑: 赵晓丽)

Evaluation of STEM Teaching Based on Classroom Observation: Protocols, Elements and Methods

ZHAN Zehui & LV Siyuan

(School of Information Technology in Education, South China Normal University, Guangzhou 510631, China)

Abstract: *As an important way of evidence-based evaluation, classroom observation can provide direct and reliable evidence and measure teaching practice of STEM education. Through content analysis and comparative study of international STEM classroom observation protocols, we explore the components and practice characteristics to provide reference for the design and implementation of STEM teaching evaluation in China. The study found that the protocols focus on the evaluation of teaching strategies, teaching content, teaching implementation and teaching evaluation with prior knowledge, cooperative learning, and scientific inquiry as key evaluation indicators of teaching strategies. The protocols also aim at the generating disciplinary core concepts and interdisciplinary concepts at the teaching content in addition to student participation and cognitive development for learning behaviors and possible thinking processes with conceptual thinking, critical thinking, innovative thinking and metacognitive skills as cognitive development goals. Attention is paid to the classroom culture of student-student interaction and teacher-student interaction, advocating an autonomous, democratic and interactive classroom atmosphere. For application contexts, although the protocols can provide suggestions for improving teaching and learning from various aspects, there are problems such as insufficient observation of students' learning behaviors, neglect of the role of technology, and difficulty in characterizing practical and operational features. Accordingly, the study deconstructs the elements of STEM classroom observation around four key dimensions: student learning, teaching, classroom culture, and technology application; and constructs a three-dimensional matrix of STEM classroom evaluation based on three aspects: evaluation design, evidence collection, and evaluation implementation, which provides references to understand STEM classrooms, and implement STEM teaching evaluation.*

Key words: *STEM education; teaching evaluation; classroom observation; teacher professional development*