

师范生“人在环中”虚拟教学实训效果及实训设计策略研究——基于 39 项实验与准实验元分析

Human-in-the-Loop Simulations Within Extended Reality for Preservice Teacher Training:

A Meta-Analysis and Instructional Design Guidelines

李园园¹, 李奕霏², 顾小清², 王龚^{1*}

¹ 上海师范大学, 分析测试与超算中心

² 华东师范大学, 教育学部, 教育信息技术学系

* swan@shnu.edu.cn

【摘要】 “人在环中”虚拟教学实训在师范生培养中日益受到关注, 但其效果与设计策略尚缺系统证据。本研究采用元分析方法, 整合 2013—2025 年 4 月间 Web of Science、Scopus 及 Google Scholar 收录的 39 篇实证文献。结果显示, 该类实训对师范生培养具有中等程度正向促进作用 ($g=0.535$)。调节效应分析表明, 规划阶段应重视实训场所、模拟学生自主性、任务多样性与实训时长; 实施阶段需完善实训介绍、理论准备、预演练习、即时提示与训后反馈等教学训练支持的设计; 评估阶段未发现显著调节效应。研究据此提出相关设计策略建议。

【关键词】 虚拟教学实训; 师范生培养; “人在环中”仿真; 元分析

Abstract: Human-in-the-loop simulations within extended reality environment show growing promise in preservice teacher education, yet evidence on effects and design remains limited. This meta-analysis synthesized 39 studies (2013–2025) from Web of Science, Scopus, and Google Scholar. Results indicated a moderate positive effect on training outcomes ($g = 0.535$). Moderator analyses showed that, in the planning phase, training effectiveness was significantly influenced by training setting, simulated student autonomy, task diversity, and simulation duration. In the implementation phase, instructional supports such as training introduction, theoretical preparation, rehearsal, prompts, post-training feedback, and other scaffolds played a significant role, whereas no significant moderating effects were identified in the evaluation phase. The study further offered design implications for optimizing such simulations.

Keywords: Extended reality, Simulation-based learning, Preservice teacher training, Human-in-the-loop simulation, Meta-analysis

1. 前言

教学活动具有高度复杂性与动态性, 教师需在真实课堂中持续应对学生差异与情境不确定性, 这对师范生在职前阶段获得充分且多样化的教学实践提出了更高要求。尽管基于真实课堂的见习与实习被普遍视为师范生培养的重要途径, 但在实施过程中仍面临个性化指导不足、实践情境与任务类型受限, 以及试错成本与伦理风险较高等现实困境 (高文心, 2025)。为回应上述挑战, 虚拟教学仿真逐渐被引入师范生培养情境, 并成为近年来教师教育研究的重要方向。随着技术演进, 虚拟仿真由早期的案例文本、角色扮演与非交互式动画, 发展为融合 VR、MR 与 AR 等扩展现实 (Extended Reality, XR) 技术的沉浸式实训环境, 在交互性、

真实感与情境复杂性方面显著提升。然而,已有研究指出,缺乏真实师生互动的 XR 仿真难以充分再现课堂教学的核心机制,限制了师范生在师生对话与课堂应变等关键教学能力上的发展 (Levin, 2024)。基于此,“人在环中”的虚拟教学仿真 (Human-in-the-Loop Simulation, HITLS) 逐渐兴起,通过引入可与人实时互动且具备一定自主反应能力的模拟学生 (Simulated Student),更贴近课堂教学的动态与不确定特征 (Li et al., 2025)。尽管 HITLS 被认为具有重要应用潜力,但现有实证研究对其实训效果的结论尚不一致,且已有研究表明,技术引入本身并不能保证教学成效,其效果在很大程度上取决于实训设计策略的合理性。然而,目前鲜有研究对 HITLS 实训策略进行系统的实证检验与分析。基于此,本研究采用元分析方法,系统评估 HITLS 对师范生培养的总影响,并进一步探究不同实训阶段设计策略的调节作用,以期为该领域的实践提供可靠的实证依据,并为高效能实训方案的设计提供策略启示。

2. 研究综述

2.1 基于 XR 的“人在环中”虚拟教学仿真系统类型与特征

现有面向师范生培养的 HITLS 主要可分为两类:以真人实时操控为核心的数字木偶式系统 (Avatar-Mediated System),如 Mursion (Hudson et al., 2019),以及以计算机技术驱动的智能体型系统 (Agent-Mediated System),如 MetaClass (Zheng et al., 2025)。前者中,模拟学生的外形与基础动作由计算机生成,而语言互动与行为反应由真人演员实时操控,在互动自然性、情境应变与对话真实感方面高度贴近真实课堂,但其高度依赖人力投入,通常需要脚本设计与排练,限制了大规模应用。后者则完全由计算机系统驱动,通过规则模型、认知模型或生成式人工智能模拟学生的认知、行为与情绪反应,在可扩展性、自主性与个性化模拟方面具有显著优势,尤其随着大语言模型的发展,其在自然语言互动、学习行为多样性与情感表达方面的表现显著提升。HITLS 可部署于 AR、VR 与 MR 等不同 XR 环境中,在沉浸水平、交互方式与系统配置上存在差异,进而影响师范生的临场感、参与度与训练体验。多数系统在环境呈现与学生角色形象上追求较高真实度,但高真实度是否必然转化为更优教学成效仍存在争议。同时,不同系统在模拟学生行为自主性、情绪表达与活动复杂性方面差异显著,从依赖人工操控或预设脚本,到具备记忆与生成能力的智能体均有涉及。总体而言,既有研究主要从技术层面呈现 HITLS 的多样化形态,但不同系统类型及其关键特征如何转化为可解释的师范生培养成效,仍缺乏系统性的实证整合与比较分析。

2.2 既有综述研究的结论与不足

现有综述研究为理解 HITLS 在教师教育中的应用提供了初步基础,但整体仍较为分散。多数综述聚焦于单一 XR 类型,其中以 MR 研究最为常见 (Dieker et al., 2023),而 VR 与 AR 相关综述相对较少。相关研究普遍肯定 XR 仿真在促进教师专业发展方面的积极作用,同时也指出技术稳定性不足、交互体验不自然等共性问题。仅有少数综述明确聚焦“人在环中”这一交互范式,且研究对象主要限于 MR 环境下的数字木偶式系统 (Bondie et al., 2021),尚未覆盖智能体型系统及其他 XR 环境类型。在实训设计层面,既有综述多停留在对单一设计要素的描述性总结,缺乏系统性的理论整合与有效性检验。总体而言,现有研究以定性综述为主,尚未通过元分析量化评估 HITLS 对师范生培养的整体效应,也未在统计层面检验不同系统类型与实训设计策略的调节作用,从而限制了对 HITLS 支持师范生培养机制的深入理解。

3. 编码框架的构建

为系统探究有效教学设计的理论依据,本研究以 NLN Jeffries 模拟理论作为元分析的编码框架 (Jeffries, 2021)。该理论已在医学模拟教育领域得到广泛应用,其“情境—背景—设计—

仿真体验—结果”五要素模型为体验式、互动式、协作式及学习者中心的仿真活动设计提供了系统指导。本研究在保留五要素结构的基础上，从师范生实训流程视角对编码维度进行阶段化整合，构建了“规划—实施—评价”三个阶段与五要素相对应的分析框架（见图1）。

3.1 实训规划阶段

3.1.1 情境 (Context) & 背景 (Background)

(1) 实训场所：仿真环境的真实性直接影响师范生的临场感与信任建立。本研究将实训场所编码为“教室”、“其他线下场所”（如实验室、研讨室）、“在线会议室”、“混合式场所”。

(2) 课程整合：探讨虚拟实训是否融入师范生课程并计入学分，编码为“是”或“否”。

3.1.3 设计 (Design)

技术设计维度：(3) XR 类型：编码为“混合现实 (MR)”“虚拟现实 (VR)”或“增强现实 (AR)”；(4) 沉浸水平：编码为“非沉浸式”“半沉浸式”或“全沉浸式” (Huang 等, 2021)；

(5) 交互方式：编码为“自然交互”（如语言、手势）、“非自然交互”（如菜单操作）或“二者结合”；(6) HITLS 类型：编码为“数字木偶式系统”或“智能体式系统”；(7) 模拟学生自主性：依据模拟学生控制方式，编码为“完全受控”“半自主”与“完全自主”。“完全受控”指行为与情绪完全由真人操控或预设脚本驱动；“半自主”指语言或关键决策由真人或规则控制、非语言行为由基础算法生成；“完全自主”指由人工智能系统独立生成行为、情绪与决策。

任务设计维度：(8) 个性化实训：是否根据师范生个体需求对训练任务进行调整，编码为“是”或“否”；(9) 任务类型：编码为“分布式练习”或“整合式练习”。其中，“分布式练习”指将训练任务分散于多个时段实施，以间隔练习方式促进技能保持与迁移；“整合式练习”指在单次训练中整合多项教学技能或教学环节，强调综合运用与整体迁移；(10) 任务多样性：是否在训练中设置多种不同类型的任务，编码为“是”或“否”；(11) 实训时长：依据单次模拟训练持续时间，编码为“30 分钟及以下”或“30 分钟以上” (Wu 等, 2024)；(12) 重复训练：是否包含多轮或重复实施的模拟训练，编码为“是”或“否”；(13) 训练方式：编码为“个体训练”或“协作训练”。

3.2 实训实施阶段

3.2.1 模拟体验 (Simulation Experience)

参与人员角色维度：(14) 实训引导者：编码为“高校教师”“研究人员”“高校教师与研究人员”或“高校教师与技术人员”；(15) 师范生：根据是否包含观察性学习，将师范生参与角色编码为“参与者”“观察者”或“参与者&观察者”。

实训支持维度：在训练前阶段，对 (16) 实训介绍、(17) 理论准备和 (18) 预演练习分别编码为“是”或“否”；在训练中阶段，对是否提供 (19) 即时提示进行编码，编码为“是”或“否”；在训练后阶段，对 (20) 反思活动、(21) 训后反馈以及 (22) 其他支架支持（如总结性复盘讨论）分别编码为“是”或“否”。

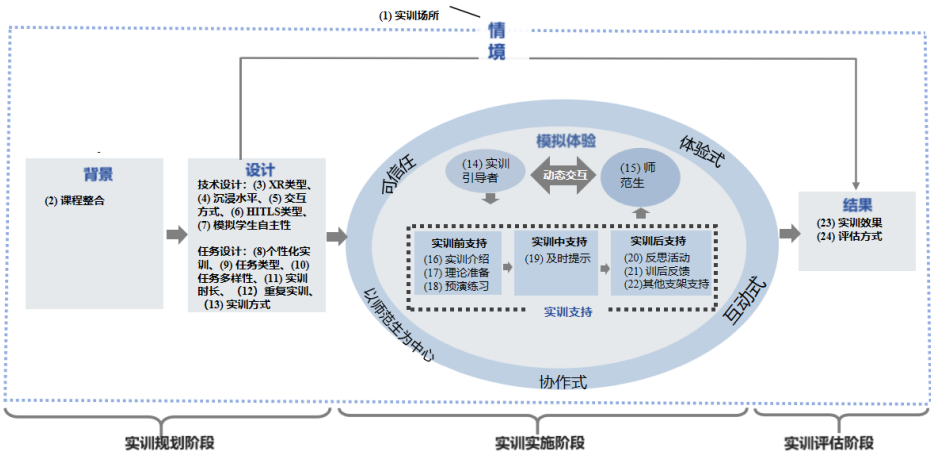


图 1 基于 NLN Jeffries 模拟理论的编码框架

3.2 实训评估阶段

3.2.1 结果 (Outcome)

(23) 训练结果类型：依据 Kirkpatrick 四层次培训评估模型 (Kirkpatrick & Kirkpatrick, 2016)，将训练结果编码为“满意度”“态度”“知识”或“技能”；(24) 评价时点则根据结果测量发生的时间，将评价方式编码为“训练中”或“训练后”。

4. 研究设计

本研究采用元分析方法对相关实证研究进行整合分析。研究过程严格遵循 PRISMA 报告规范 (Page et al., 2021)，以提升研究筛选、编码与分析过程的透明性与可重复性。

4.1 文献检索与筛选

本研究以 Web of Science、Scopus 和 Google Scholar 为数据来源，系统检索 XR 环境下“人在环中” (HITLS) 应用于师范生培养的实证研究。已有研究表明，联合使用上述数据库可在文献覆盖范围与结果稳定性之间取得较优平衡 (Harzing & Alakangas, 2016)。检索词基于数据库主题词表与代表性文献反向追溯确定，并采用布尔逻辑组合。关键词分为两组：一组为师范生相关术语 (如 preservice teacher、student teacher、teacher candidate、teacher trainee)，另一组为 XR 与虚拟仿真相关术语 (如 VR、AR、MR、XR、metaverse、simulation、virtual human、avatar、agent)，组内以“OR”连接，组间以“AND”连接。

鉴于沉浸式虚拟现实技术在教育领域的系统研究始于 2013 年 Oculus Rift 发布，本研究将检索时间范围设定始于 2013 年，于 2025 年 4 月完成检索，初步获得文献 9,478 篇。为确保筛选过程的可靠性，两名研究者依据统一标准独立筛选并交叉核验。文献纳入标准包括：(1) 2013—2025 年发表的同行评审文献；(2) 研究对象为师范生，且明确采用 XR 环境下的 HITLS；(3) 实验或准实验研究设计；(4) 提供计算效应量所需统计信息；(5) 同一研究重复发表保留一篇。

初筛剔除非英文文献及重复文献后剩余 4,277 篇；进一步排除主题不符或研究设计不合要求的文献，保留 54 篇；因统计信息不足剔除 15 篇，最终纳入 39 篇文献，共提取 142 个效应量用于元分析。

4.2 文献编码与数据分析方法

本研究以 NLN Jeffries 模拟理论为编码框架 (见图 1)，由两名研究者独立完成文献编码，编码一致性良好 (Cohen’s Kappa 均 > 0.90)。统计分析采用 Comprehensive Meta-Analysis

(CMA) 3.0 软件。效应量基于实验组与对照组的均值差异计算；当原始研究未报告均值与标准差时，则依据样本量、t 值或效应方向进行估算。所有效应量统一转换为 Hedges'g，以降低小样本偏倚。研究间异质性通过 Cochran's Q 检验与 I² 指标进行评估，并依据 Higgins 等（2003）的标准判断异质性水平。在此基础上，进一步开展调节变量分析，以检验不同系统特征与教学设计要素对效应量差异的解释作用。

4.3 发表偏倚检验与异质性检验

为检验发表偏倚，本研究采用漏斗图进行评估（见图 2）。结果显示效应值分布总体对称，仅有少量离群点，未见明显发表偏倚。Egger 回归检验结果不显著（ $p=0.212$ ）。此外，Classic fail-safe N 为 4375，显著高于 $5k+10$ 阈值（即 145），表明本研究元分析结果具有较高稳健性。异质性检验结果显示，Q 统计量为 1507.737（ $p<0.001$ ）， $I^2=90.648\%$ （ $>50\%$ ），表明研究间存在显著异质性。因此，本研究采用随机效应模型对总体效应值进行估计。此外，效应量在不同研究间差异较大，这说明可能存在调节变量对实训效果产生影响，这为后续进行调节效应分析提供了依据。

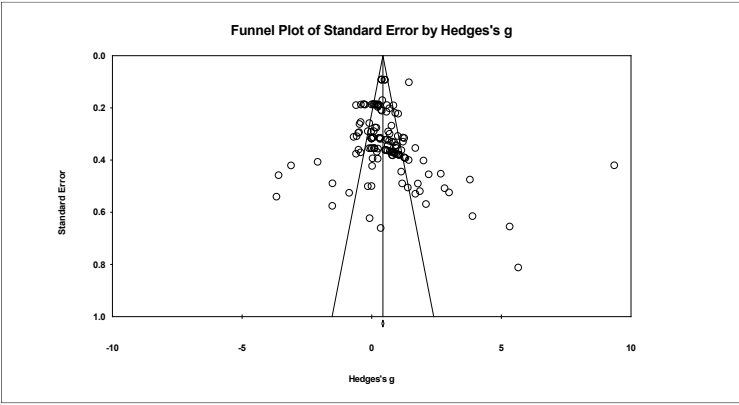


图 2 效应值分布漏斗图

5.研究分析

5.1 实训效果分析

经分析，纳入研究的效应值分布在 -3.665 至 9.363，其中 39 项为负向效应。总体合并效应值为 $g=0.535$ （ $p<0.001$ ，95% CI [0.392, 0.678]），属于中等效应水平（Cohen, 0.2 小，0.5 中，0.8 大），这表明 HITLS 对师范生教学能力具有稳定且显著的正向促进作用。为探究 HITLS 对不同实训结果维度的影响，本研究对“满意度”“态度”“知识”“技能”进行分组分析（见表 2）。组间异质性检验显示差异不显著（ $Q_B=0.052$ ， $p=0.997$ ），这表明 HITLS 对各维度影响整体一致。具体而言，效应值均为正向，知识提升最大（ $g=0.635$ ），随后为满意度（ $g=0.524$ ）、技能（ $g=0.521$ ）与态度（ $g=0.510$ ），这表明 HITLS 在师范生的认知、技能与情感层面的实训效果均具有积极作用。

表 1 整体实训效果分析结果

效应数	效应值	95% 置信区间		双尾检验		异质性检验			
		下 限	上 限	Z 值	p 值	Q 值	χ ² f 值	p 值	I ²

142	0.5	0.	0	7	.	150	]	.	90.6
35	392	.678	.338	000	7.737	41	000	48	

表 2 不同训练效果维度的效应分析

实训效果	效 应数	g	z	95% 置 信区间	Q _B	df	p
					0.052	3	0.997
满意度	39	0.524	5.712	[0.344, 0.703]			
态度	51	0.510	3.713	[0.241, 0.779]			
教学知识	8	0.635	1.171	[-0.427, 1.696]			
教学技能	44	0.521	4.408	[0.289, 0.752]			

5.2 实训设计策略

为探究如何设计有效的实训策略以优化 HITLS 对师范生的培养效果，本研究依据 NLN Jeffries 模拟理论框架，分别从规划、实施与评估三个阶段对实训设计策略的调节效应进行分析。

5.2.1 实训规划阶段的实训设计策略

在情境与背景设计方面，“实训场所”呈显著调节效应（ $Q_B=9.087, p<0.05$ ），不同场所效应量依次为混合式场所 > 其他线下场所 > 在线会议室 > 教室，表明虚拟仿真与真实空间结合的混合式环境更能增强师范生的临场感与情境信任。相比之下，“课程整合”虽未显著，但纳入正式课程并计入学分的研究平均效应量较高，提示 HITLS 系统性融入师范生课程可能有助于提升实训效果。

在技术与任务设计方面，“模拟学生自主性”（ $Q_B=26.398, p<0.05$ ）、“任务多样性”（ $Q_B=6.232, p<0.05$ ）与“实训时长”（ $Q_B=5.750, p<0.05$ ）均呈显著调节效应。具体而言，半自主型模拟学生的 HITLS 系统效果最佳，其次为完全受控型系统，完全自主型系统相对较弱；设置多样化任务以及单次训练超过 30 分钟亦有助于提升实训成效。

此外，虽然“XR 类型”“沉浸水平”“交互方式”“HITLS 类型”“个性化实训”“任务类型”“重复训练”和“训练方式”等变量未达显著水平，其效应量分布仍呈现趋势：不同 XR 类型均表现稳定正向效果，非沉浸式略高于半沉浸式和全沉浸式；数字木偶式系统整体优于智能体式系统；任务定制、分布式练习、协作训练以及多轮训练的设计在总体上表现更优。

5.2.2 实训实施阶段的教学设计策略

在参与角色维度上，“实训引导者”与“师范生”未呈显著调节效应，但效应量存在差异。结果显示，同时担任参与者与观察者的师范生，其实训成效优于单一角色，这提示结合实践操作与观察性学习的参与方式更有利于教学技能发展。

在实训支持设计方面，多项支持要素在实训前、实训中及实训后阶段表现出显著调节效应。在实训前阶段，“实训介绍”（ $Q_B=7.044, df=1, p=0.008<0.05$ ）、“理论准备”（ $Q_B=3.934, df=1, p=0.047<0.05$ ）与“预演练习”（ $Q_B=19.746, df=1, p<0.05$ ）均对实训效果产生显著促进作用，表明在正式模拟前为师范生提供清晰说明、理论支撑与预演机会，有助于其更有

效地投入模拟教学情境。在实训中阶段,“即时提示”(Q_B = 16.889, $df = 1, p < 0.05$)表现出显著调节效应,说明在模拟过程中适时提供教学提示,有助于支持师范生进行教学决策与课堂应对。在实训后阶段,“训后反馈”(Q_B = 8.414, $df = 1, p = 0.004 < 0.05$)与“其他支架支持”(Q_B = 7.502, $df = 1, p = 0.006 < 0.05$)对实训成效具有显著正向影响,而“反思活动”未呈现显著调节效应(Q_B = 0.251, $df = 1, p = 0.616 > 0.05$)。这一结果表明,在当前 HITLS 实训研究中,结构化的反馈与支架支持可能比开放式反思活动更直接地促进实训效果。

5.2.3 实训评价阶段的教学设计策略

在实训评价阶段,“训练结果类型”与“评价方式”均未表现出显著调节效应。无论是满意度、态度、知识还是技能维度,实训效果均呈现出稳定的正向影响。同时,从效应量分布来看,在实训过程中即时开展评价的研究,其效果略优于实训结束后进行评价的研究,但整体差异未达到显著水平。

6. 讨论与结论

6.1 研究结论

(1) 基于 XR 的 HITLS 实训对师范生教学实训具有显著的中等促进效应($g = 0.535$)。这表明,基于 XR 的 HITLS 能够在总体上有效促进师范生教学能力的发展。不同训练结果维度(满意度、态度、知识与技能)之间的效应量差异不显著,这说明 HITLS 对师范生教学发展的支持具有一定的整体性特征,有助于其认知、情感与行为层面的协同提升。

(2) HITLS 的实训效果并非由技术形态本身决定,而受关键设计策略的显著调节。HITLS 的实训效果主要受关键设计策略而非技术形态本身调节。实训规划阶段的调节分析显示,实训场景、模拟学生自主性、任务多样性及模拟时长显著影响成效,其中混合式场景、半自主型虚拟学生、多样化任务以及单次训练较长均有利于提升效果;而 XR 类型、沉浸水平、交互方式及系统类型未表现显著作用,说明单纯提升技术沉浸度或更换系统形态并不必然改善实训成效。

(3) 实训支持设计在 HITLS 实训实施过程中发挥关键作用。实训介绍、理论准备、预演练习、即时提示及训后反馈与其他支架支持均显著增强实训效果;虽师范生参与角色未显著调节效应,但参与与观察结合的方式显示一定优势,表明 HITLS 的有效实施依赖多阶段、结构化的实训支持。

(4) 当前 HITLS 研究在评价层面的设计仍较为有限。不同结果维度(知识、技能、态度、满意度)及不同评估时机对整体效应的影响差异不显著,表明当前评估方式可能未能充分捕捉到实训效果的情境敏感性与复杂性,这暗示了未来研究需开发更具鉴别力的评估框架。

6.2 HITLS 设计策略建议

(1) 在实训规划阶段,应优先关注教学情境与任务结构设计。研究表明,相较于单一线上或线下环境,混合式实训场景更有利于发挥 HITLS 的训练优势。因此,在条件允许的情况下,应将 XR 支持的 HITLS 与真实教学空间有机融合。同时,建议采用半自主型虚拟学生设计,在保持情境真实性的同时,避免过度复杂的生成行为对师范生造成认知负担。此外,通过设置多样化任务并适当延长单次模拟时长,有助于师范生在更充分的练习中逐步整合教学知识与教学技能。

(2) 在实训实施阶段,应强化分阶段、可调控的训练支持机制。实训前阶段应系统设计实训说明、理论准备与预演练习,帮助师范生建立清晰的任务认知框架;实训过程中,应通过即时提示支持其教学决策与行为调整;实训后阶段,则应通过训后反馈与结构化支架支持,

促进经验整合与能力提升。相较于形式化的反思活动,这类嵌入式支持机制在当前 HITLS 实训中表现出更直接促进作用。

(3) 在实训评价阶段,应拓展评价维度与时点设计。尽管现有研究未发现评价方式的显著调节效应,但结果显示,在实训过程中开展评价较事后评价更具潜在优势。未来 HITLS 实训设计可进一步结合过程性评价与情境化表现评价,关注师范生教学行为的迁移效果,以提升评价结果对教学改进的解释力。

综上所述,本研究通过元分析系统考察了 HITLS 在师范生教学实训中的应用效果,结果表明该模式整体上对师范生教学能力发展具有显著促进作用,但其成效更依赖于教学情境、任务结构与训练支持等设计策略,而非技术系统本身。然而,本研究仍存在一定局限,如纳入实证研究数量有限、部分调节变量样本不均衡、实训结果测量方式差异较大的等。未来研究有必要在扩大样本基础的同时,进一步细化设计变量与评价指标,强化理论引导,深入探索 HITLS 在支持师范生教学能力发展的内在机制与实施路径。

参考文献

- 高文心.(2025).高水平师范大学的实践性课程与师范生社会化.高教发展与评估,41(06),79-86+133.
- Levin, O. (2024). Simulation as a pedagogical model for deep learning in teacher education. *Teaching and Teacher Education*, 143, 104571. <https://doi.org/10.1016/j.tate.2024.104571>
- Li, M., Pan, C., Zhang, W., Ling, J., Sun, J., Liu, X., ... & Zhu, J. (2025). Enhancing pre-service teachers training with AI-simulated students: providing realistic classroom experiences without impacting real learners. *Interactive Learning Environments*, 1-18.<https://doi.org/10.1080/10494820.2025.2591238>
- Hudson, M. E., Voytecki, K. S., Owens, T. L., & Zhang, G. (2019). Preservice teacher experiences implementing classroom management practices through mixed-reality simulations. *Rural Special Education Quarterly*, 38(2), 79–94. <https://doi.org/10.1177/8756870519841421>
- Zheng, L., Jiang, F., Gu, X., Li, Y., Wang, G., & Zhang, H. (2025). Teaching via LLM-enhanced simulations: Authenticity and barriers to suspension of disbelief. *The Internet and Higher Education*, 65, 100990.
- Dieker, L., Hughes, C., & Hynes, M. (2023). The Past, the Present, and the Future of the Evolution of Mixed Reality in Teacher Education. *Education Sciences*, 13(11), 1070. <https://doi.org/10.3390/educsci13111070>
- Bondie, R., Mancenido, Z., & Dede, C. (2021). Interaction principles for digital puppeteering to promote teacher learning. *Journal of Research on Technology in Education*, 53(1), 107–123. <https://doi.org/10.1080/15391523.2020.1823284>
- Jeffries, P. R., & National League For Nursing. (2016). *The NLN jeffries simulation theory*. Wolters Kluwer.
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's four levels of training evaluation. Association for Talent Development.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology*, 134(134), 103–112. <https://doi.org/10.1016/j.jclinepi.2021.02.003>

- Harzing, A.-W., & Alakangas, S. (2015). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>