

语言教学中技术赋能形成性评估的系统性综述

A Systematic Review of Technology-Enabled Formative Assessment in Language Teaching

王艾琳¹, 郑春萍^{2*}, 张华阳³¹² 北京邮电大学 ³ 深圳职业技术大学* amilo@163.com

【摘要】 本研究系统回顾了2012-2024年技术赋能形成性评估应用于语言教学的116项实证研究，总结了此类评估在语言教学中的基本特征与应用。对学习管理系统、多模态评估平台、基于游戏化的评估平台、虚拟世界与仿真评估平台等八类评估平台进行分类，并分析具体特点及其在即时反馈、个性化学习等方面的有效性。研究表明，技术赋能的形成性评估能显著提升学生参与度、学习动机和语言能力，在写作和口语提升上效果突出。而且，其有效性与教师技术素养、学生自我调节能力及评估设计合理性紧密相关。本文总结了该领域的优势与挑战，为未来教学实践提供了有益参考。

【关键词】 形成性评估；语言教学；技术赋能；系统性综述

Abstract: In recent years, research on technology-enhanced formative assessment in language teaching has thrived. This article reviewed 116 empirical studies (2012-2023), analyzing 8 categories of assessment platforms and tools like LMS, ITS, e-portfolio systems, AWE, etc., and their effectiveness. It found that such assessment boosts students' engagement, motivation, and language proficiency, especially in writing and speaking. Also, its effectiveness depends on teachers' teaching skills, students' self-regulation, and assessment design. This paper summarizes the advantages and challenges of technology-enhanced formative assessment, offering insights for language teaching.

Keywords: Formative assessment, Language teaching, Technology-enhanced, Systematic review

1. 研究回顾

1.1. 形成性评估定义和发展

形成性评估（Formative Assessment）最早由斯克里文（Scriven, 1967）提出，并由布鲁姆（Bloom, 1969）引入教学评价领域。其核心理念是通过持续的反馈与调整，促进学习者在教学过程中的发展。根据不同研究视角，形成性评估可以划分为过程说与实践说。过程说强调评估是教师收集学习证据，解释证据，提供反馈，以促进学习者不断进步的动态循环。实践说则侧重于评估的课堂操作，强调通过具体的工具和策略，如自评和互评，促进学习者在互动中反思和改进语言能力。两种学说的结合，有助于教师更全面地理解形成性评估的应用，特别是在技术赋能的背景下。

1.2. 技术赋能的形成性评估与语言教学

近年，技术赋能形成性评估成为教育领域的热门话题，大量实证研究展示了技术的使用方式，以及它们如何促进学习者语言能力的提高。例如，学习管理系统（LMS）如Moodle和Blackboard被广泛用于在线课堂的评估，提供集中化的评估与反馈机制（Panadero & Jonsson, 2013）。在写作方面，自动写作评估工具（AWE）如Turnitin和Grammarly已被证明在提高学生写作能力方面具有显著效果（Jiang & Yu, 2020）。而阅读技能方面，交互式计算机化评估系统（InCAS）被证实可以用于识别儿童阅读能力的强项和弱项并追踪其阅读能力（Merrell

& Tymms, 2007)。基于游戏化的评估平台如 Kahoot!, 不仅能提高学生的参与度, 还能通过互动性的任务提升学习动机 (Kim & Ryu, 2013)。研究还探讨了在线语言学习项目中的评估及电子档案评估 (Baturay & Daloğlu, 2010)。

尽管如此, 现有研究仍主要集中在具体工具的应用上, 而对技术如何整体融入语言教学中, 尤其是其如何促进形成性评估的循环过程和互动性方面的讨论较少。近年来, 越来越多的研究开始关注如何将技术工具与语言教学目标紧密结合, 以促进学生的深度学习和技能提升。

因此, 本文的研究旨在通过系统梳理近年来的实证研究, 回答以下三个核心问题: (1) 相关研究在文献计量方面呈现出怎样的特征? (2) 相关研究在实证研究方面呈现出怎样的特征? (3) 技术赋能的形成性评价应用于语言教学有什么优势与挑战?

2. 研究设计

本研究严格遵循 PRISMA 分析框架开展系统性综述 (Aromataris & Pearson 2014; Harris et al. 2014)。初步选定 19 本 SSCI 索引 I 区和 2 本 SSCI 索引 II 区, 共 21 本国际权威学术期刊, 限定发表时间为 2012-2024 年, 共得到文献 2777 篇。随后, 通过精确筛选和全文阅读后, 剔除文献 2659 篇。最终, 根据相关关键词筛选确定文献共计 116 篇, 如图 1。针对筛选获得的 116 项实证研究, 借鉴系统性综述编码方法进行编码 (郑春萍等 2021), 并运用文献计量软件 bibliometrix 进行可视化分析。

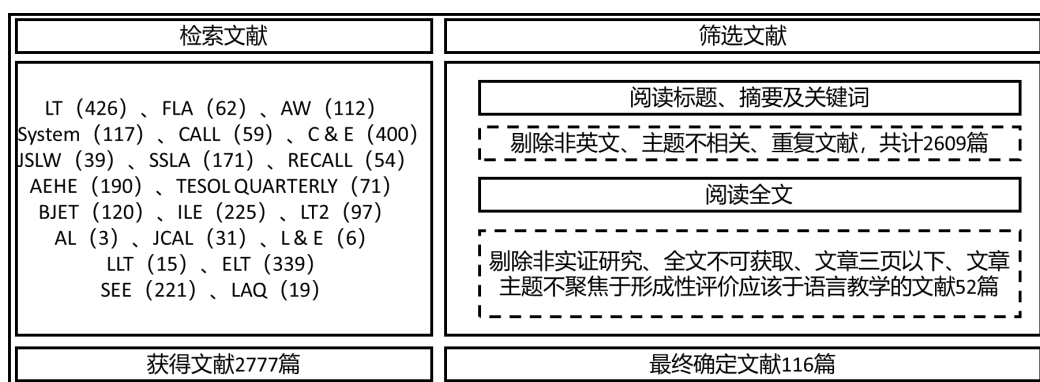


图 1 依据 PRISMA 相关标准的文献检索与筛查过程图

3. 文献发表的基本特征

3.1. 发表期刊、年份与合作机制

在期刊发文方面, 近年来 (2012-2024), 本研究所选国际期刊共刊载相关实证研究成果 116 项, 此领域仍属前沿话题。其中, 刊发相关成果最多期刊为 Assessing Writing (31 项), 其次为 CALL (20 项), System (19 项)。发文最多年份为 2023 (21 项), 2022 (17 项), 其次为 2021 (15 项)。从文献引用看, 有 8 篇被引超 100 次, 有 28 篇被引超 50 次, 有 43 篇被引超 30 次, 有 68 篇被引超 10 次。

在合作机制方面, 本研究发现第一作者单位共来自 23 个国家和地区, 其中, 亚洲 (57 项, 49.14%) 和北美洲 (39 项, 34.78%) 的研究成果占比较大, 其次是欧洲 (10 项, 8.62%), 大洋洲 (5 项, 4.31%), 南美洲 (4 项, 3.45%)。在通讯作者单位方面, 分别是中国 (38 项, 32.8%), 美国 (33 项, 28.4%) 等, 多国家单位合作发文方式仍然较少。

3.2. 理论架构

在 116 项研究中, 仅有 21 项明确提出了理论基础, 部分研究者基于技术赋能的视角引入了数字化学习理论、多媒体学习理论和 Vygotsky (1978, 1987) 的社会文化理论 (8 项)。

此外，从心理学视角出发，研究者应用了社会认知理论、自我决定理论、建构主义理论、归因理论、项目反应理论、活动理论、概化理论等；从语言学视角出发，研究也探讨了基于技术的二语习得理论和写作认知理论。可见，技术赋能下的形成性评估应用于语言教学涉及跨学科的理论体系，尽管已有一定理论依据，但在如何具体应用和深入挖掘技术对学习者认知发展的影响方面，仍有较大的提升空间。

3.3. 教育层次、语言与课程类型

研究对象涵盖了多个教育层次，包括学龄前阶段（2 项）、小学阶段（18 项）、初高中阶段（17 项）、本科阶段（68 项）、研究生阶段（4 项）及成人阶段（12 项）。而学龄前和研究生阶段的研究较少。此外，技术赋能的形成性评估在不同语言背景下也有显著应用，尤其是英语为二语的学习者（88 项），其母语涉及汉语、韩语、马来语、印地语等。研究还关注了不同语言技能的教学，尤其是在写作（63 项）、综合技能（29 项）和口语（17 项）方面。尽管如此，研究在阅读（3 项）、听力（3 项）和翻译（1 项）方面较少，这些领域的技术应用仍然是未来的研究重点。

4. 技术赋能的形成性评估平台与工具的有效性分析

技术赋能的形成性评估通过多种类型的在线平台与工具实现，具有数据驱动、即时反馈、多模态支持和游戏化设计等特性。对近十年实证研究的梳理发现，有 8 类典型的形成性评估平台或工具被应用于语言教学。本文依据 Golonka et al. (2014) 关于技术支持语言学习的分类方法，对相关平台和工具进行了筛选与补充。

4.1. 学习管理系统 (LMS)

学习管理系统 (LMS) 是支持教师组织课程、管理资源和监控学习进展的在线平台 (He & Wang 2024)。典型平台包括 Moodle、Blackboard 和 Edulastic 等。

LMS 的核心优势在于促进持续参与、多层次反馈和个性化学习 (Fernando, 2020)。具体而言，LMS 平台如 Moodle 通过多样化的评估工具（如在线测试、小组讨论、同伴互评等），促进了学生在语言学习中的深度参与和学习效果 (Fernando, 2020)。此外，LMS 内嵌的测评工具能够根据学生表现自动调整难度，并结合学生的行为数据（如学习时长、作业提交、论坛参与、视频观看、互动行为等），通过学习者建模进一步实现个性化教学。

然而，研究也发现 LMS 平台在支持形成性评估方面的局限。首先，平台自动反馈的精准度和深度尚未达到要求，对于复杂的语言技能（如写作和口语），反馈机制仍依赖教师的人工评分。同样地，教师反馈通常局限于评分，而非深入的个性化指导 (He et al., 2024)。此外，部分 LMS 平台在设计上存在不合理之处。不同于 Moodle 将一周所需的所有学习元素和工具连续放置在一个直观的大纲结构中，一些平台（如 Blackboard）将任务元素分布在多个页面中，这增加了认知负担，并可能影响学生的任务参与度 (Rubin et al., 2010)。LMS 平台在支持同伴互评方面能力也有限。平台内嵌的测验、论坛和小组讨论等功能缺乏有效的整合，导致实时反馈和互动协作难以无缝衔接，影响形成性评估的整体效果。

4.2. 智能教学系统 (ITS)

智能教学系统 (Intelligent Tutoring Systems, ITS) 是一类通过人工智能和大数据技术，实时分析学习者行为并提供个性化反馈的教育平台。ITS 能够根据学生的认知水平和学习表现，为语言学习中的复杂技能提供定制化支持 (Haridas et al., 2020)。表 1 为几种典型 ITS 平台。

表 1 支持语言教学形成性评估的典型智能教学系统及其功能特点

代表工具	研发团队	特点及功能
AmritaITS	印度阿姆里塔	模拟与专业教育工作者的一对一互动，根据学生的认知技能、

	大学	知识库量身定制教学材料，并根据交互日志预测学习成绩。
Writing-Pal	哈佛创始团队（助梦空间）	专注写作策略指导，通过分阶段教学模块（预写、起草、修订）提供写作策略支持。
iSTART	美国亚利桑那州立大学	专注于阅读理解技能，通过交互式训练模块与即时反馈提升学生的批判性阅读和元认知能力。

智能教学系统在语言教学形成性评估中的关键优势在于其量身定制的交互式学习、自适应教学路径的动态调整，以及对复杂技能评估的针对性支持（Haridas et al., 2020）。例如，协同写作系统 Writing-Pal（W-Pal）作为一种 ITS，通过动画教学代理和交互式小游戏，引导学生在写作的预写、起草与修订过程中掌握策略，构建了以反馈与修订循环为核心的写作教学模式（Allen et al., 2019）。相比传统 LMS，W-Pal 更注重策略教学而非简单的任务管理或目标追踪。对于阅读技能，iSTART 系统结合交互式训练与即时反馈机制，提升学生的批判性阅读能力和元认知水平（McNamara et al., 2015）。在听力与口语技能方面，英语口语自动评估工具（如 CAES）通过语音识别技术提供即时反馈，帮助学生改善发音与流利度。未来，ITS 有望通过更精细的学习者建模和更复杂的算法，融合多模态交互方式，并与虚拟现实等教育技术平台集成，为评估提供更加个性化与交互式的方案。

4.3. 电子档案系统 (e-Portfolio System)

电子档案系统（e-Portfolio System）在 2000-2010 年间广泛使用，通过整合多模态资源，支持学习者记录、展示和反思学习过程与成果，如表 2。

表 2 支持语言教学形成性评估的典型电子档案系统及其功能特点

代表工具	研发团队	特点及功能
Mahara	新西兰的 Catalyst IT 公司	提供基于 DigCompEdu 框架的作品集，学生可以插入多模态资源，如音视频、ppt、甚至社交媒体帖子或网页等。
LinguaFolio	美国国家外语中心	侧重目标设定和进度追踪，支持多语言环境下的学习成果展示与反思（Ziegler 2012）。

研究显示，学习者可以将结合多种媒体类型的内容（如文本、图像、绘画、照片、插图、演示文稿、文档、数据文件、录音、社交媒体），将帖子、博客文章、自拍照、视频博客等内容添加到他们的电子文件夹中，展示学习过程和成果。电子档案系统的多模态特性，有助于深度学习的实施，尤其是在认知学习层面，而视觉与言语渠道的结合已被证明有效促进学术学习（Mayer, 2017）。近年的实证研究普遍认可电子档案袋对自我调节学习的作用，特别是在生成性与反思性学习策略的实施上。然而，研究也指出，当前一些平台的设计和操作简单性可能对技术素养较低的学生构成障碍，同时其反馈机制的质量和交互性仍有改进空间（Lam, 2022）。

4.4. 自动写作评估工具

自动写作评估工具（Automated Writing Evaluation, AWE）是通过语料库、自然语言处理和机器学习技术实现的一类智能评估工具，能够根据词汇复杂性、句法多样性、语篇结构、语法使用、单词选择和内容开发来分析文本的复杂程度。典型的 AWE 系统如下表 3 所示。

表 3 支持语言教学形成性评估的典型自动写作评估工具及其功能特点

代表工具	研发团队	特点及功能
批改网	中国某公司	提供整体评分与评语，对词汇、句子、篇章结构、内容打分。
eRevise	美国匹兹堡大学	提供不同颗粒度的反馈，基于自然语言处理提供自动评分。
Grammarly	乌克兰某公司	提供反馈，包括正确性(语法和写作技巧)、清晰性(简洁和可读性)、投入性(词汇和多样性)和传递性(正式性、礼貌和自信)。

Criterion® 美国教育服务考试中心 提供分数和诊断性反馈，如语法、用法等方面错误类型判断。

由上表可知，自动写作评估工具（AWE）的核心优势在于能够为学习者提供即时同步的反馈，帮助他们快速识别写作中的问题（Shintani, 2016）。除了细致的错误反馈，AWE 工具还能提供个性化的诊断和自助访问资源（如网络词典、词库、编辑工具、学生作品集等），支持学生的写作改进（Chong, 2017）。现有研究普遍验证了 AWE 工具在词汇、句法和语法层面的有效性。然而，研究也指出，AWE 工具的反馈通常集中于写作的表面特征（如拼写、用词和语法），而对内容开发关注不足，未能反映写作的多模态、上下文和社会方面（Vojak et al., 2011）。例如，Liu 和 Kunnan（2016）的研究发现，WriteToLearn 在处理复杂语言现象（如冠词和介词使用）时表现欠佳。类似地，Lavolette, Polio 和 Kahng（2015）指出，Criterion 在识别复杂语法错误时的正确率仅为 75%。在语篇层面，Liao（2016）指出，AWE 工具在宏观层面，特别是在篇章结构和内容发展上，反馈往往过于笼统且缺乏功能性建议。此外，大多数 AWE 工具的功能集中于编辑和修改阶段，对写作全过程（如预写、计划和起草阶段）的支持较为有限，仅有少数工具能够提供宏观层面段落逻辑和文本结构的反馈（Fernando, 2018）。而在学习者能动性方面，计算机生成的反馈可能限制学生的表达选择，削弱其能动性，并导致学生更多地迎合自动评价标准，而非积极探索写作的创造性表达（Zaini, 2018）。

4.5. 基于游戏化的评估平台

基于游戏化的形成性评估平台通过融入积分、任务、奖励等游戏元素，能够唤起如成就感和兴奋感的人类情感，并通过互动和反馈机制，促进语言学习的持续进步，典型平台如表 4。表 4 支持语言教学形成性评估的基于游戏化的评估平台及其功能特点

代表工具	研发团队	特点及功能
Kahoot!	挪威某教育科技公司	教师通过设计测验、调查和讨论，创建游戏社区互动以刺激学生参与。学生通过 PIN 码加入游戏，回答问题并获得积分。
Quizizz	印度某教育科技公司	具有独特的实时互动和家庭作业评估模块。教师可创建测验、学生以自己的节奏作答，系统实时显示得分和排名。
Socrative	加拿大某教育技术公司	允许教师开发个性化活动，如简答题、测验、投票并即时反馈。学生可参与个人或小组任务，提交匿名答案，甚至参加友好竞争。
Duolingo	美国某教育技术公司	提供短小的课程，结合游戏化元素（如每日学习目标、虚拟货币奖励、闯关任务）帮助学生在游戏体验中学习语言。

研究表明，游戏化评估平台能在短期内提升语言学习效果。例如，Duolingo 通过积分和进度条提高词汇和语法习得效率（Shortt et al., 2023），角色扮演与任务闯关增强写作和沟通能力（Guo et al., 2024）。此外，游戏化对情感和社交互动有积极影响，如积分奖励可降低焦虑、提升学习动力（Zhang & Hasim, 2023），Kahoot! 的排行榜机制则促进同伴互动并激发竞争意识。

然而，尽管短期内游戏化评估表现出明显的积极效果，其长期效果目前仍然褒贬不一，未能有效弥补平台在深度学习、社交互动以及个性化反馈方面的缺陷。多项研究指出 Duolingo 等平台的问题：在学习内容上，Dehghanzadeh（2020）及 Loewen 等人（2019）发现其多为浅层次重复练习、简单翻译，侧重接受性技能训练，忽视生产性技能培养。社交互动上，Marques-Schafer 等人（2018）指出其讨论论坛互评机制缺失。此外，Duolingo 采用的行为主义方法未考虑语言学习的意义和文化背景，Teske（2017）认为其缺乏实用或文化技能相关活动，难以满足社会文化理论角度的教学需求。在元认知层面，部分研究显示学生对游戏化工具的兴趣会随时间减弱，导致学习动机下降（Loewen et al., 2019）。过度依赖竞争元素会削

弱学生内在动机。如 Marques-Schafer 和 da Silva Orlando (2018)、Huynh 等 (2016) 发现, Duolingo 因任务重复、缺乏挑战性, 使部分学生失去兴趣甚至停止使用。

4.6. 多模态评估平台、虚拟世界与仿真评估平台、生成式人工智能

多模态与仿真评估平台在技术赋能的形成性评估中发挥重要作用。多模态平台整合视觉、听觉和文本数据, 结合深度学习技术, 实现细粒度反馈。例如, Zheng et al. (2024) 开发的演讲评估系统可实时分析语言、表情和动作, 实现闭环反馈。视频反馈因涵盖语音、语气、肢体语言等非语言特征, 在可理解性和情感互动方面更具优势 (Yiğit & Seferoğlu, 2021)。然而, 其推广受限于高技术门槛, 目前仍依赖自主研发的实验性平台。虚拟与仿真评估则通过沉浸式交互环境提升语用能力, 如 Second Life 通过任务导向设计评估商务交流, 而仿真平台在航空、医学英语等职业语言训练中展现独特价值。然而, 高昂成本、教师技术素养和课程整合仍是推广难点, 未来应聚焦优化技术、降低门槛, 并探索与人工智能的融合。

生成式人工智能 (Generative AI) 正在重塑语言评估。ChatGPT、BERT、Kimi 等工具基于自然语言生成技术, 能够实时生成内容、模拟互动、定制评估材料, 并提供纠错、示例与建议, 促进创造性思维。其依赖 Transformer 架构 (如 GPT-4), 通过多头自注意力机制建模语境, 并结合预训练与微调, 支持多维度评估。然而, 该技术在低资源语言和跨文化语境中的适应性仍有限, 反馈可能缺乏语境适切性, 甚至引入误导信息, 同时隐私与数据伦理问题亦待解决。未来应提升其语境适应能力、增强数据透明度, 并探索与智能评估技术的融合, 以构建更精准、高效的语言评估体系。

5. 结语

技术赋能的形成性评估正重塑语言教学评估体系, 为实时监测学习表现、优化教学策略提供新路径。本研究基于 2012-2024 年 116 项实证研究的系统性综述, 发现该领域正快速发展, 研究多聚焦本科生, 采用定性、定量及混合方法, 数据收集多样, 分析方式匹配研究问题。理论上, 多借鉴社会文化理论, 并结合教育学、语言学、心理学及数据科学, 为不同文化、学龄及语言技能的评估提供支撑。技术赋能的形成性评估在语言教学中的应用优势主要体现在以下几个方面: (1) 实时反馈与优化, 智能反馈机制减少传统评估的滞后性, 助力学习策略调整 (Butler et al., 2021); (2) 提升教学成效, 多模态数据整合提高评估精准度, 互动工具增强学习体验 (Gikandi et al., 2011); (3) 新兴技术支持, 机器学习、虚拟现实等助力个性化评估与沉浸式学习 (He & Wang, 2024)。但也仍存挑战: (1) 理论与实践整合, 技术评估如何与传统方式协同仍存争议 (Can Daşkın & Hatipoğlu, 2019); (2) 技术局限与反馈效用, 自动评分在写作逻辑、语篇组织及口语非语言特征分析上不足, 学生反馈接受度存在差异 (McNamara et al., 2015; Zhang & Hyland, 2022); (3) 教师适应与资源不足, 部分教师缺乏培训, 学校资源亦受限 (Schildkamp et al., 2020)。基于以上分析, 本研究建议: (1) 深化应用研究, 关注学生个体差异; (2) 加强理论与实践结合, 提高可操作性; (3) 推动跨学科合作, 开发更具实用性与创新性的评估工具, 助力语言学习全面发展。

参考文献

- 郑春萍, 徐畅, 张娴, 刘涵泳. (2021). 近十年数字游戏应用于语言教学的系统性综述. 现代教育技术, 31(06), 41-48.
- Aromataris, E., & Pearson, A. (2014). The systematic review: An overview. *AJN The American Journal of Nursing*, 114(3), 53-58.

- Allen, L. K., & Perret, C. A. (2016). Commercialized writing systems. In D. S. McNamara & S. A. Crossley (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 145–162). NY: Taylor & Francis, Routledge.
- Baturay, M. H., & Daloğlu, A. (2010). E-portfolio assessment in an online English language course. *Computer Assisted Language Learning*, 23(5), 413–428.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. *Teachers College Record*, 70(10), 26-50.
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455.
- Can Daşkın, N., & Hatipoğlu, Ç. (2019). Reference to a past learning event as a practice of informal formative assessment in L2 classroom interaction. *Language Testing*, 36(4), 513-537.
- Chong, I. (2017). How students' ability levels influence the relevance and accuracy of their feedback to peers. A case study. *Assessing Writing*, 31, 13–23.
- Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaei, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, 34(7), 934–957.
- Fernando, W. (2020). Moodle quizzes and their usability for formative assessment of academic writing. *Assessing Writing*, 46, 100485.
- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57, 2333–2351.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70-105.
- Guo, K., Zhong, Y., Zainuddin, Z., & Chu, S. K. W. (2024). Applying game - related methods in the writing classroom: A scoping review. *Education and Information Technologies*, 29(4), 4481–4504.
- Haridas, M., Gutjahr, G., Raman, R., Ramaraju, R., & Nedungadi, P. (2020). Predicting school performance and early risk of failure from an intelligent tutoring system. *Education and Information Technologies*.
- Harris, J. D., Quatman, C. E., Manring, M. M., Siston, R. A., & Flanigan, D. C. (2014). How to write a systematic review. *The American Journal of Sports Medicine*, 42(11), 2761-2768.
- He, M., & Wang, L. (2024). Implementing assessment as learning in online EFL writing classes. *RELC Journal*.
- Huynh, D., Zuo, L., & Iida, H. (2016). Analyzing gamification of “Duolingo” with focus on its course structure. In *International Conference on Games and Learning Alliance* (pp. 268–277). Cham: Springer.
- Jiang, L., & Yu, S. (2020). Appropriating automated feedback in L2 writing: Experiences of Chinese EFL student writers. *Computer Assisted Language Learning*, 35(7), 1329-1353
- Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks. *Educational Technology Research and Development*, 61, 549–561.

- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2), 50–68.
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1-18.
- Liu, S., & Kunnan, A. J. (2016). Investigating the application of automated writing evaluation to Chinese undergraduate English majors: A case study of WriteToLearn. *CALICO Journal*, 33(1), 71–91.
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311.
- Marques-Schafer, G., & da Silva Orlando, A. A. (2018). Languages learning conceptions and Duolingo: A critical analysis on its proposals and learners' experiences. *Texto Livre - Linguagem e Tecnologia*, 11(3), 228–251.
- Mayer, R. E. (2017). Using multimedia for e - learning. *Journal of Computer Assisted Learning*, 33(5), 403–423.
- McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35–59.
- Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer - adaptive assessments. *Journal of Computer Assisted Learning*, 23, 27–35.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.
- Rubin, B., Fernandes, R., Avgerinou, M. D., & Moore, J. (2010). The effect of learning management systems on student and faculty outcomes. *Internet and Higher Education*, 13, 82–83.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. RAND McNally.
- Schildkamp, K., Van Der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602.
- Shintani, N. (2016). The effects of computer - mediated synchronous and asynchronous direct corrective feedback on writing: A case study. *Computer Assisted Language Learning*, 29(3), 517–538.
- Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile - assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, 36(3), 517-554.
- Stevenson, M., & Phakiti, A. (2014). The effects of computer - generated feedback on the quality of writing. *Assessing Writing*, 19, 51-65.
- Teske, K. (2017). Duolingo. *CALICO Journal*, 34(3), 393–401.
- Vojak, C., Kline, S., Cope, B., McCarthy, S., & Kalantzis, M. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition*, 28(2), 97–111.
- Yiğit, M. F., & Seferoğlu, S. S. (2021). Investigating the effect of video - based feedback on perceived feedback quality. *Pamukkale University Journal of Education*, 51, 92–122.

- Zaini, A. (2018). Word processors as monarchs: Computer - generated feedback can exercise power over and influence EAL learners' identity representations. *Computers & Education*, 120, 112–126.
- Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, 13, 1030790.
- Zhang, Z. (V.), & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, 100586.
- Zheng, C., Zhang, T., Chen, X., Zhang, H., Wan, J., & Wu, B. (2024). Assessing learners' English public speaking anxiety with multimodal deep learning technologies. *Computer Assisted Language Learning*.
- Ziegler, N. A., & Moeller, A. J. (2012). Increasing self - regulated learning through the LinguaFolio. *Foreign Language Annals*, 45(3), 330–348.