

数字赋能英语读写教学评价： 概念、路径和理据

张 莉

东华大学外语学院，上海

收稿日期：2025年5月12日；录用日期：2025年6月27日；发布日期：2025年7月9日

摘 要

数字化时代和人工智能时代带给高等教育的发展既有挑战也有机遇。公共英语的教学评一体化改革应顺应时代潮流积极拥抱信息技术。该研究探讨智慧教育模式中数字赋能的教学评价，包括数字评价的构念(知识能力组成元素)、数字评价实施的内外环境(信息化一代的师生与数字测评的平台)，数字形成性评价和数字终结性评价等研究问题。论文以东华大学公共英语课程数字转型即智慧教改为研究案例，提出新时代的语言测评应融合物理空间、社会空间和数字空间，构建以数据驱动、数字赋能，以核心素养培养为方向的智慧教学评价新模式。除了剖析英语读写评价的痛点重点和智慧评价的路径与理据，该研究也对立德树人素养价值引领的评价提出一些建议。

关键词

数字赋能，教学评价，英语读写，数字化智慧化教育

A Probe of Digital-Empowered Assessment on Teaching English Reading into Writing: Concept, Methods and Rationales

Li Zhang

College of Foreign Languages, Donghua University, Shanghai

Received: May 12th, 2025; accepted: Jun. 27th, 2025; published: Jul. 9th, 2025

Abstract

The digital age and the era of artificial intelligence present both challenges and opportunities for the development of higher education. The integration of teaching and evaluation reform for General

English should embrace information technology in accordance with the trends of the times. This study explores the digital empowerment of teaching evaluation in smart education models, including the constructs of digital evaluation (elements of knowledge and ability), the internal and external environments for implementing digital evaluation (teachers and students of the information age and the platforms for digital assessment), as well as research issues such as digital formative and summative evaluations. The paper uses the digital transition of the integrated English reading into writing at Donghua University as a case study, proposing that language assessment in the new era should integrate physical, social, and digital spaces to create a new model of intelligent teaching evaluation. This model is data-driven, digitally empowered, and focuses on cultivating core competencies. In addition to analyzing digital-empowered English reading and writing assessment, as well as the pathways and rationale for intelligent evaluation, this research also offers some recommendations for evaluation on moral education.

Keywords

Digital-Empowered, Teaching Assessment, English Reading into Writing, Digital and Intelligent Education

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. 研究背景和研究问题的提出

人类已经进入了数字化时代和人工智能时代。“数”是数字化，是数字赋能教育全过程，是互联网大数据驱动教学，是区块链融入教育评价[1]；“智”是智慧教育、智慧课堂与智慧校园，是人工智能、AI 教师和虚拟仿真课程，是新基建背景人工智能技术融入高等教育。数智时代教育的本质没有发生根本改变，依然是立德树人的根本任务，依然是培养德智体美劳全面发展的社会主义建设者和接班人这样的教育方针。党的二十大提出深入实施科教兴国战略，首次将教育、科技、人才三大战略整体部署，一体推进[2]。新时代背景带来了高等教育的新环境新业态和诸多挑战，对其学科建设、课程设置、评估方式、教师队伍等等均产生巨大影响[3]。

《地平线报告》(Horizon Report)是全球教育科技发展的一个风向标，可为教师带来前沿的教育趋势分析。该报告指出未来高等教育五大趋势，其中就有技术方面的趋势，未来教育是基于人工智能的教学创新[4]。报告介绍了将对高等教育教学产生重大影响的六项关键技术，其中和语言评价密切相关的有应用于个性化学习的预测性人工智能(AI-Enabled Applications for Predictive, Personal Learning)和生成式人工智能(Generative A)。其应用分别对应语言测评中的分级测试(placement test)、诊断性考试(diagnosis test)、学业型考试(achievement test)、过程性评价(formative evaluation)和终结性评价(summative evaluation)等测评类型。

“智慧教育”从目的而言是指通过构建技术融合的学习环境，让教师能够施展高效的教学方法，让学习者能够获得适宜的个性化学习服务和美好的发展体验，使其由不能变为可能，由小能变为大能，从而培养具有良好的价值取向、较强的行动能力、较好的思维品质、较深的创造潜能的人才。“智慧教育系统”包括现代化的教育制度、现代化的教师制度、信息化一代的学生、智慧学习环境及智慧教学模式五大要素。其中，智慧的教学模式是整个智慧教育系统的核心组成[5]。语言教学的智慧测评是语言智慧教育模式中很重要的一部分。智慧测评的概念如何界定，智慧测评的内外环境要求，智慧评价的具体类

别及实施过程都是值得研究的学术命题。

2020年,中国发布《21世纪核心素养5C模型研究报告(中文版)》。原有“4C”基础上提出的“5C”核心素养包含跨文化交际力(Culture Competency)、创新力(Creativity)、思辨力(Critical thinking)、合作力(Collaboration)、沟通力(Communication)。素养模型拥有五个一级维度和16个二级维度[6]。数智时代培养人才的核心素养在5C核心素养的基础上,更倾向培养学习者的数字素养、提出问题的能力和在人工智能协助下的学习能力[7]。《中国智慧教育蓝皮书(2022)》立足“智慧教育发展处于起步阶段”的客观实际,探索建立了由4个一级维度、12个二级维度构成的评价指标体系[8]。数智时代下,这些素质培养目标对创新教学团队的智慧教学实践指明了方向,对智慧测评的构念(测什么)和路径(怎么测)提出了导向性要求。

基于以上研究背景和前人的研究,本文着力探讨智慧语言教育模式中的数字赋能语言教学评价,主要包含下面三组研究问题:

- 1) 智慧语言教学评价测什么?是语言知识、语言能力,还是价值、素养,有哪些构念?
- 2) 智慧语言评价怎么实施?内外环境的要求是什么?具体测试的样态有哪些?
- 3) 如何平衡智慧语言教学测评中的数字形成性评价和数字终结性评价?

接下来论文将结合公共英语读写课程的智慧教学与评价实践,分别给出具体的分析和回答。

2. 智慧语言教学评价:概念、路径及理据

智慧语言教学评价是语言智慧教育模式中包含的重要内容之一。通过智能技术开展的评价既可以通过实时数据收集和统计分析对教学过程进行评价和反拨驱动再设计;又可以通过阶段性过程评价和终结性评价对教学目标达成与否和教学效果进行评价。基于复合型构念的语言智能测评往往在智慧测试平台或自适应学习平台上进行,因而对该研究命题的探讨也使得智慧学习和测试平台的研发和建设变得十分迫切。

2.1. 智慧语言教学评价的构念

《教育与心理测量标准》(Standards for Educational and Psychological Testing,以下简称《标准》)[9]对construct“构念”的定义是“测试所测量的概念(concept)或特性(characteristic)”(p.11)。这一抽象概念考虑为何而测和测量的是什么[10]。智慧语言教学评价则是把语言测试的构念置于信息化数字化的技术应用环境,其构念主体内容仍旧以语言测试构念的特质/能力观视角和任务/情景观视角为主,特别对于英语读写模块的教学内容而言,交互观的测试构念并不是英语读写侧重的方面[11]。举例来说,英语读写授课往往依托(数字或新形态)教材中的文本阅读模态分析来引导学习者掌握阅读的相关知识和能力,通过设计的产出学习任务比如写作练习来检测学习者的书面语表达能力。“5C”核心素养均为软能力,它们需要复合进读写语言能力和读写情景任务的完成过程中加以评价。比如跨文化交际力的运用效果可以编入来自阅读文本中设计的思考分析题目里,通过学习者的分析回答进行测试(示例于本论文第3部分微格教学评价活动);又比如思辨力可以通过显性的英语表达语句 which means that 的量化数据,基本划分等号学习者是否运用了批判性思辨能力。

复合型构念的语言测评是指测试的内容是综合了多维多项的内容,并不是单一测试语言知识或语言能力,而是将若干语言教学的具体能力(比如阅读能力中的上下文猜词义、构词法知识、语篇主题理解和相关学科背景知识调用辅助理解等阅读微技能)和“5C”核心素养中的跨文化交际力、思辨力进行复合。“十八大”以来在教育界推行的混合式教学与评价新导向还要求语言教学与测评能自然融入体现课程思政立德树人的价值引领内容。因此,复合型语言测试构念呈现复杂、动态、多元的特点。在智

慧教育模式下,智慧测试的构念(测什么)依然是复合型语言测试的构念,并没有因为“智能化”技术手段的加入发生本质上的改变。在原有复合型构念的测试内容融合基础上,数智时代的智慧测评还新增融入向人工智能比如 GPT 或 Deep Seek 提出学习问题的能力以及人机协作完成学习任务和拓展实践个性化学习的数字素养等内容。总之,智能化的数字测评的构念主体仍然是语言综合知识和能力,以及复合融入的核心素养和数字素养能力。智慧教育测评以教学过程全程数据实时提供和科学统计分析反拨重构教育过程的方式,赋能、转型语言的教学过程和教学效果评价向信息化、智能化的语言教育和语言测评发展转变。

2.2. 智慧语言测试的实施环境和样态

近三年的高校学习者是 00 后伴随智能手机出生和使用的一代人。2023 年的教学观察中,他们在教室里使用平板、智能手机查阅电子教材、使用生成式人工智能技术进行语言自学和拓展学习的现象开始出现并逐渐变得多了起来。2025 年年初 Deep Seek 的横空出世使得人工智能赋能教学和智慧课程、智慧校园的建设话题讨论变得更加广泛而热烈。种种变化说明智慧教育系统中信息化一代的学生和教师已经到位,实施智慧语言测试的内部环境也已基本成形。同时,教师和学生的数字素养作为学术研究热点之一也处在高度关注中。

智慧语言测试的外部环境是指智慧校园、智慧教室、智慧测试云平台等“硬件”条件。华东地区是实践智慧教育的头阵,上海肩负着“排头兵”的重任。东华大学在 2023 年下半年拓展了原有的智慧教室范围,将之扩大到两个校区的主要教学楼教室全覆盖。目前这些智慧教室的数字赋能教学功能主要是智慧云录播系统,可以同步教师在教室的传统面授,进行网络直播和点播回看。笔者还了解到浙江某高校的智慧教室功能更加全面,可以以全息技术异地同时开展合作创新教学,真正促进中西部教学联动,教育机会公平和优质教育资源共享等学习共同体的教育战略部署和落地实施。语言测试方面,外语智慧测试平台主要是国内两大外语出版巨头——外研社和外教社——在研发、带动、服务全国的外语智慧教学。以外研在线 unipus 旗下的 i-test 智慧测试云平台为例,目前功能已经涵盖机器组卷、AI 评卷包括评阅主观题型翻译和作文,提供测试数据反馈和统计分析等等方面。平台以其数字智能技术极大节省了教师人工出卷和阅卷的时间,优化了混合式教学过程和测评。

智慧语言测试的具体样态呈现进一步多样的发展态势,不仅仅是语言单一或复合的构念只通过文字表述的样态进行测评,还可以考虑声音、图画、动图、视频等多模态的形式进行学习任务和测试内容的设计与评价。除了单一模态向多模态转变,智慧语言测评还能运用大数据技术支持 portfolio 电子档案袋的形式搜集和分析动态学习数据实施增值评价[12]。智慧平台上搭建知识图谱、能力图谱,推荐学习者个人学习路径。AIGC 智能体伴学引导学习者关注自身学习成长,引导其进行基于数据的自我成长和终身学习。这样智慧测试的样态具有了由静态向动态的转变,由指向外的他适应转向指向内的自适应。除了智慧测试平台的机评,智慧教育实施过程中教师还可以融合使用教师人工评、组织学习者开展自评、引导学习者进行互评,外加上智慧教室的大量过程性教学评估数据,智慧教育能描绘学习者学习画像,分析学习者学习轨迹,提供动态调整学习的策略和建议,真正体现“以学生发展为中心;以学生学习为中心;以学习效果为中心”新三中心的变革落地。

2.3. 语言智慧过程性评价 vs 语言智慧终结性评价

“四新”建设的核心要求是学科交叉融合培育拔尖创新复合人才;数字化的教育特点是技术赋能差异化教学促进个人成才。智能化个性化自适应终身学习则是不久的未来可预见的智能教育的新样态。就目前语言教学而言,像公共英语类课基本仍沿用大规模班级授课(比如:2023 年秋季学期一门《英语 3》

读写课有 26 个平行班学生人数 1200 左右, 教学团队有六名教师); 期末终结性评价时仍旧是工业化时期的大规模标准化期末测试卷统一测试和标准化统一阅卷(出卷和阅卷的格式要求甚至呈现“模板化”、“固化”的苗头)。不可否认标准化统一测评在某一历史阶段有其优越性, 可站在新时代智慧教育的大门口, 语言测评若还一味延续传统, 没能积极迎接数字化智能化的技术趋势带来的势不可挡, 没能积极迎接数字化智能化的技术趋势带来的势不可挡, “显然就不太适宜, 没有做到与时俱进, 顺势而为”。现阶段的语言教学测评提出的是过程性评价和总结性评价相结合的方式, 注重过程性评价。然而, 从测试学的角度来说, 真正的过程性评价的举措是不会再有期末考试终结性评价结合使用的。课程的全部测评数据均为教学过程中采集, 包括学习者的出勤、平时作业练习、课堂学习活动的参与完成度、学习者的自评和互评, 学习任务的机评和师评等等, 按一定权重综合评定学习者学完一门课的整体学习效果和表现。

该研究认为, 语言的智慧测评不妨设计为语言的智慧过程性评价和增值评价。增值评价采用诸如 portfolio 电子档案袋等形式和区块链等技术手段将学习过程数据去中心化、全程采集, 科学计算, 分析得出学习者的学习效果出具可信安全的评价数据分析报告, 按照人文底蕴、科学精神、学会学习、健康生活、责任担当、实践创新等综合素质的增幅大小评定等级或分数。因为站在终身学习、自主学习的智慧教育角度, 语言测试是不可能具有真正意义的终结性评价的。任何一张符合测试构念拥有较好测试信度和效度的试卷, 也都只能是阶段性的学习效果测试(assessment for learning), 测评目的是促进学习, 改善学习, 终身学习。AI 自适应学习的崛起使得人工智能驱动的自适应学习平台日益受到关注和应用。智慧测试平台和智慧学习平台可以进一步研发拓展功能、协同共建, 优化学习者的学习路径, 提升学习效果和改善学习体验[13]。促进学习的(智慧)评价应体现在(智慧)教学的全过程[14]。

3. 智慧“教、学、评”一体设计：以大学英语读写《英语 3》跨文化思辨微格教学为例

公共英语读写课程拥有配套的智慧测试云平台, 比如《英语 3》读写综合课采用外研社的 i-test 智慧测试云平台, 依托该平台 AI 组卷并开展两次阶段(过程性评价)能力测试和期末试卷组卷(终结性评价)。课程教学周期内的不同类型测试均围绕测试学习者的英语阅读和英语写作的知识点掌握, 英语阅读和写作的能力运用。课程测试内容呈现复合型测试构念, 主要以知识和能力复合为主。“5C”核心素养软能力以及数字学习的素养在课程组规定动作的“统一测试”计划实施中不是很明显。素养目标和高阶能力主要依靠课程团队教师个人在授课环节当中进行具体设计和教学实施。智慧“教、学、评”一体教学设计选取作者 2023 年秋所授《英语 3》班级的微格学习任务设计和教学实施, 示例了教学单元“童年记忆”里的课文原句改编设计为学生个人思考题。班级每位学习者在校本 SPOC 平台(泛雅超星学习通)上传自己的理解思考和主观回答, 在教师引导下完成异步讨论。教师在课堂面授环节对部分学习者的思考回答样本进行点评讲解, 发现学习者掌握的不足进而设计后续的延展式语言教学或补救式教学。图 1 是教师对某份学生线上个人思考回答之后的点评讲解示例。

这样的微格教学设计在具有“两性一度”的学习任务引导完成中促进学习者的语言知识学习和语言能力运用, 培养其高阶思维能力和核心素养, 同时自然融入思政育人元素。通过优秀作业/练习的示例点评, 既能重构和丰富教学内容, 以点带面示范性教学激发学习者的学习动力(向优秀同伴看齐), 又能引导学习者发现自身学习的不足(比如回答时没有结合例子或因为忽略了语气加强词 extremely 和模糊限定词 possibly 回答得特别绝对), 实施基于学习者实际表现的差异性/拓展型教学和补救式教学, 体现了“因材施教”和“以学习者为中心”理念的贯彻落地。

基于课文原句的思考题改编如下:

Seen through the eyes of Chinese youngsters, do you label many “normal” Western childcare practices as extremely bizarre and possibly harmful to children? Please use examples to explain your judgment.



Figure 1. A sample of teacher-annotated learners' individual thinking and upload
图 1. 学生线上思考作答的教师点评示例

该课例并入单元的思政教学教案设计，获得了 2024 年度全国高校外语课程思政教学案例大赛二等奖。2024 年起作者加入了第三期产出导向法云共同体暨教育部多语种教研虚拟教研室，继续对该课(准)智慧教学进行打磨，丰富教研(评价)数据，进行数据分析和三角验证。

4. 反思和总结

数字化智能化时代呼唤差异性、个性化、智能化的教育新模式的研究和建设。其中，语言智慧测试是语言智慧教育模式中很重要的组成部分。本文以英语读写智慧教创改革举措举例，阐述了智慧语言测试的构念仍以语言复合能力测试内容为主，兼顾数字素养、核心素养的增值评价；梳理了智慧语言测试的实施内外环境要求，条件已基本具备，在进一步完善功能推进智慧教室、智慧校园的建设中。论文还认为，智能语言测评应强化过程评价，探索增值评价，真正做到大数据驱动的形成性评价。人工智能背景下的高等语言教育不能丢失“温度”，人工智能作用于教学和评价的“程度”要努力做到匹配和“平衡”[15]。基线测评重视每位学习者的学习起点和个体差异；过程监控利用信息化、智能化的大数据，绘制学习者在人文底蕴、科学精神、学会学习、健康上火、责任担当、实践创新六方面变化的学习者成长画像；终结测评需改变单一测试语言知识和能力的旧模式，多采用多元方式实施形成性评价，展现学习者六个发展维度的变化幅度；增值分析则是大数据可视化对比分析学生的学习“起点”与学习“结果”，客观、公平、科学、准确地对每一位学习者以及教师、学校进行教育投入和教育产出的评价分析。

自然融入语言教学的素养育人和价值引领评价往往是教学创新路上的堵点问题。对此，本文基于智慧语言教育测评分析提出的注重过程性评价和形成性评价，采用 portfolio 电子档案袋进行增值评价，尝试描绘生成的学习者学习路径，或许能成为解决智慧教育评价堵点问题的一种尝试。

致 谢

本文撰写得到了东华大学课程教学创新和教学能力提升培育项目——《英语 3》的支持。项目号 114-03-0007014。作者也感谢第三期产出导向法云共同体暨教育部多语种教研虚拟教研室对该研究的指导。

参考文献

- [1] 全立新. 区块链在数字教学资源评价中的应用[J]. 中国现代教育装备, 2022(7): 9-12.
- [2] 陈廷柱, 管辉. 教育数字化: 转型还是赋能[J]. 中国远程教育, 2023, 43(6): 11-18.
- [3] 中华人民共和国教育部. 推进教育数字化[EB/OL].
http://www.moe.gov.cn/jyb_xwfb/s5148/202304/t20230427_1057446.html, 2023-04-27.
- [4] 肯学君. 2024 即来: 深度解读教学创新权威报告, 看见教创赛与一流课程的未来[EB/OL].
<https://mp.weixin.qq.com/s/yPL3Zp7wPkUQgXA4LvpFXQ>, 2023-12-13.
- [5] 知乎专栏. 智慧教育打造个性化教学、互动性教学环境[EB/OL]. 2023-05-30.
<https://zhuanlan.zhihu.com/p/633341591>, 2023-09-18.
- [6] 魏锐, 刘坚, 白新文, 等. “21 世纪核心素养 5C 模型”研究设计[J]. 华东师范大学学报(教育科学版), 2020, 38(2): 20-28.
- [7] 中华人民共和国教育部. 《中国智慧教育蓝皮书(2022)》: 学生数字素养培养成效显著[EB/OL].
http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2023/2023_zt01/mtbd/202302/t20230213_1044392.html, 2023-02-13.
- [8] 蒋里. AI 驱动教育改革: ChatGPT/GPT 的影响及展望[J]. 华东师范大学学报(教育科学版), 2023, 41(7): 143-150.
- [9] AERA, APA and NCME (2014) Standard for Educational and Psychological Testing. American Educational Research Association.
- [10] 金艳. 语言测试中的构念和构念效度[J]. 语言测试与评价, 2022(1): 44-57+117.
- [11] Bachman, L.F. (2007) What Is the Construct? In: *Language Testing Reconsidered*, University of Ottawa Press, 41-72.
<https://doi.org/10.2307/j.ctt1ckpccf.9>
- [12] 安富海. 探索增值评价制度 学生素质提升幅度大, 才是好学校[EB/OL].
<https://m.gmw.cn/baijia/2020-07/21/34011626.html>, 2020-07-21, 2024-01-12.
- [13] 麦可思报告. 中国—世界高等教育趋势报告(2023) [R]. 2024-01-03.
- [14] 王海啸. 从形成性评价到促进学习的评价[EB/OL]. 腾讯会议直播讲座的笔记记录(Individual Note-Taking), 2020-04-12.
- [15] 杨港, 顾世民. 人工智能时代外语教学模式变革的风险及其规避[J]. 语言教育, 2023, 11(3): 20-29.