

# 数智化背景下高校外语类听说课程多模态评价机制构建

## ——以“同声传译”课程为例

朱德红\*, 张静

枣庄学院外国语学院 山东枣庄

**【摘要】**在全球化与数智化深度融合的背景下,高校外语听说类课程的传统评价体系面临维度单一、过程性缺失、反馈滞后等结构性困境,难以全面捕捉学生的语言能力、认知策略及跨文化素养。本研究以多模态学习理论与交际语言能力模型为框架,融合人工智能、大数据等数智技术,构建“数智化多模态评价机制”。该机制涵盖多维评价内容,通过多主体协同评价模式,实现过程性数据采集与即时反馈,可为外语听说课程评价改革提供创新范式,并为推广至其他课程提出技术适配与师资培训路径。

**【关键词】**数智化;多模态评价;听说课程;人机协同

**【收稿日期】**2026年2月18日

**【出刊日期】**2026年3月31日

**【DOI】**10.12208/j.ssr.20260101

### Construction of a multimodal evaluation mechanism for college foreign language listening and speaking courses in the context of digital intelligence——A case study of the *Simultaneous Interpreting* course

Dehong Zhu\*, Jing Zhang

School of Foreign Languages, Zaozhuang University, Zaozhuang, Shandong

**【Abstract】** Against the background of globalization and digital intelligence, traditional assessment of college foreign language listening and speaking courses is limited by its single dimension, insufficient process evaluation and delayed feedback. Based on multimodal learning theory and communicative language ability model, this study constructs a digital intelligence-enabled multimodal evaluation system with the support of artificial intelligence and big data. The system adopts multi-dimensional and multi-agent collaborative assessment to realize process data collection and real-time feedback. It provides an innovative paradigm for the assessment reform of foreign language courses and offers feasible paths for wider application.

**【Keywords】** Digital intelligence; Multimodal evaluation; Listening and speaking courses; Human-machine coordination

#### 1 引言

在全球化和信息化深入发展的背景下,高校外语教育正面临前所未有的挑战与机遇。听说能力作为语言学习的核心组成部分,其评价体系亟需适应数智化时代的教育变革需求。特别是在同声传译等高端语言技能课程中,学生须具备高度的语言敏感度、快速反应能力与跨文化交际素养,而传统评价体系仍存在明显局限:一是维度偏窄,偏重语言形式准确性,忽视认知策略、非言语交际与情感态度等关键素养;二是过程缺位,重终结性评价而轻过程性反馈,难以实现对学习过程的动态支持;三是效能不足,人工评价效率低、主观

性强,技术赋能尚未形成系统性突破。

数智化技术的飞速发展破解了以上难题提供了新路径。语音识别、情感分析、机器翻译等人工智能技术,为实现对学生语言表现、翻译质量、反应速度及跨文化能力的全面、即时评价奠定了基础。因此,构建数智化驱动的多模态评价体系,不仅是提升同声传译课程教学质量的内在要求,也是推动外语教育评价现代化的重要突破口。

本研究立足于数智化背景,以同声传译课程为例,旨在构建一套融合多模态学习理论、交际语言能力模型与智能技术评价方法的综合性评价体系,理论上,拓

\*通讯作者:朱德红

展语言评价的理论边界,推动多模态评价范式的建构;实践上,设计覆盖“语言—认知—非言语—情感”四维的评价工具体系,实现教学评一体化;生态层面,积极响应《中国教育现代化 2035》中对教育评价体系改革的战略导向,探索人机协同的评价路径,为构建智能化教育评价生态提供可复制的实践案例。

## 2 理论基础

### 2.1 数智化教育评价的研究现状与趋势

随着人工智能、大数据、学习分析等技术的成熟,教育评价正经历着从“经验驱动”向“数据驱动”的范式转型。在语言教学领域,数智化评价的应用主要体现在以下几个方面:

首先,在自动化评分与反馈方面,技术已从对语法、词汇等表层语言特征的分析,发展到对口语流利度、发音准确度的评价,并逐步延伸至对写作内容逻辑与质量的评估<sup>[1]</sup>。语音识别与自然语言处理技术使得对学习者口语产出进行即时、客观的量化分析成为可能,极大地提升了评价效率<sup>[2]</sup>。

其次,在学习过程与行为分析方面,数智化技术能够捕捉传统评价难以触及的微观过程数据。例如,通过分析学生在同声传译任务中的停顿、修正、延时而副语言特征,可以间接推断其认知负荷与决策过程<sup>[3]</sup>。眼动追踪、面部表情分析等技术的引入,为进一步探索学习者的注意力分配与情感状态提供了新的视角。

然而,现有研究也存在明显局限。多数应用仍停留在对单一语言技能(如口语或写作)的自动化测评,缺乏对“语言-认知-情感”等多维能力的整合性评价框架。此外,当前,技术赋能评价多被视为对传统人工评价的“工具性替代”,如何将机器评价与教师、同伴及自我评价有机融合,构建“人机协同”的评价生态,仍是当前研究的薄弱环节。

### 2.2 多模态学习与评价的理论基础

多模态理论源于社会符号学,认为意义并非仅通过语言,而是通过图像、声音、手势、空间等多种符号资源共同构建<sup>[4]</sup>。在教育领域,多模态学习理论强调学习者通过调动多种感官通道参与学习过程,能更深入地理解和建构知识<sup>[5]</sup>。

将多模态理论引入评价领域,意味着评价视角应从单一的语言文本,扩展到学习者在完成任务过程中所调用的全部符号资源。对于同声传译这类高度复杂的认知活动而言,其表现不仅体现在译语的准确性上,还蕴含于译员的面部表情、身体姿态、语音语调等非言语信息中,这些信息共同传递着其专业性、自信度及跨

文化沟通效能<sup>[6]</sup>。因此,一个全面的评价体系必须能够捕捉并诠释这些多模态数据。

然而,多模态评价研究虽已兴起,但多集中于对多媒体作品或课堂演示的分析,对于如何系统地将多模态数据(特别是非言语与情感数据)纳入形成性评价流程,并为其建立科学、可操作的指标体系,尚缺乏深入探讨。

此外,语言测试与评价的实践始终受到特定语言能力模型的影响。从最初关注语言知识本身的结构主义模型,到强调语言交际功能的交际语言能力模型<sup>[7]</sup>,再到将策略能力、心理生理机制纳入其中的综合模型,语言能力的定义在不断拓宽。尤其是近年来,动态、复杂的语言能力观愈发受到重视。该观点认为,语言运用是认知、社会、情感因素在特定情境中动态互动的结果<sup>[8]</sup>。这为本研究提供了关键启示:对同声传译能力的评价,绝不能孤立地看待语言输出,而应将其置于一个包含认知策略、情感因素、非言语资源在内的整体框架中进行审视。

### 2.3 研究述评与本研究的理论定位

综合现有文献可知,数智化技术为教育评价带来了新的工具与方法,多模态理论拓展了评价的视野,而演进的语言能力模型则指明了评价应覆盖的维度。然而,三者之间存在明显的“割裂”状态:1)技术应用与理论深度脱节:许多技术驱动的评价研究缺乏坚实的教育学与语言学理论支撑,未能触及高端语言能力的核心内涵。2)多模态评价缺乏系统整合:现有研究多关注某一特定模态(如仅关注语音或仅关注视频),未能构建一个统整多模态数据的综合性评价体系。3)人机角色尚未厘清:在评价实践中,机器与人的职能边界模糊,如何发挥各自优势,形成合力,是亟待解决的关键问题。

本研究旨在弥合上述裂痕,通过有机融合数智化技术、多模态评价理论与现代语言能力模型,构建一个系统性的、人机协同的多模态评价框架。

### 2.4 本研究的理论框架

基于上述分析,本研究构建了如下图所示的整合性理论框架。该框架由四个相互关联的层次构成:

(1) 理论基础层:这是框架的指导思想,融合了多模态学习理论(指明评价的内容边界)、交际语言能力与动态能力观(界定评价的核心维度)以及数智化教育评价理论(提供评价的方法论)。

(2) 数智化技术层:这是框架的赋能支撑。它具体阐述了如何运用语音识别与自然语言处理技术分析语言维度的表现;利用计算机视觉与视频分析技术捕捉非言语交际行为;借助基于绩效的代理指标(如音频分析获取的停顿/填充词频率、视频分析获取的视线转

移模式等)与基于多模态数据的情感分析(如对语音副特征进行分析、对任务录像中的体态语进行观察、以及对反思文本中的情感词汇进行内容分析),分别量化和解读学生的认知策略与情感状态。

(3) 多模态评价体系层: 这是框架的核心实体。它具体化为一个由“语言-非言语-认知-情感”四个维度构成的评价内容体系,并通过机器自动评价、学生自我评价、同伴评价、教师评价四种主体协同运作。在此层

面,技术层的多模态数据被转化为各维度的具体指标,并由不同的评价主体进行解读与判断,最终形成一份整合性的评价报告。

(4) 应用目标层: 这是框架的价值归宿。构建该体系的最终目标在于: 促进教学评一体化(为教学改进提供精准依据)、赋能学生全面发展(通过多维反馈促进元认知与综合素养提升)<sup>[9]</sup>、优化教育评价生态(推动构建更加科学、公平、高效的评价环境)。

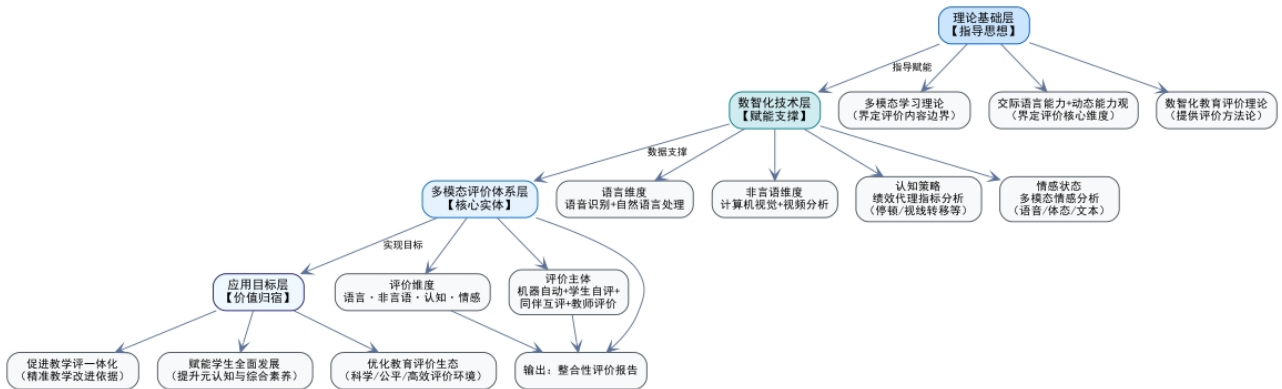


图1 数智化背景下听说课程多模态评价的理论框架

该框架清晰地勾勒了本研究从理论到实践的转化路径,为后续评价体系的具体设计与实证研究提供了清晰的蓝图。

### 3 数智化多模态评价体系的设计与构建

基于前述理论框架,本节从设计原则、核心维度、评价机制与技术路径三个层面,系统阐述数智化多模态评价体系的构建过程,形成一个科学、可操作的教学评价模型。

#### 3.1 设计原则与理念

本体系的构建遵循三大核心设计原则。首先,坚持人机协同原则,明确界定机器与人在评价活动中的角色边界与功能优势。机器擅长处理海量数据,进行客观量化分析;教师和学生则长于情境化解读、质性判断与价值引导。二者形成优势互补的评价合力,共同提升评价的效率和深度。其次,强调过程与结果融合原则,打破传统终结性评价的局限,将评价有机嵌入学习全过程。通过智能技术手段持续采集过程性数据,实现对学习者能力发展轨迹的动态追踪与即时反馈,使评价真正服务于学习改进。最后,遵循多维动态原则,该体系的设计超越单一语言技能评价的窠臼,建立涵盖语言、非言语、认知及情感等维度的综合性评价指标<sup>[10]</sup>,动态反映学生在复杂任务中展现的综合素养。

#### 3.2 核心评价维度与运作机制

本体系的核心创新在于确立了“语言-非言语-认知-情感”四个相互关联的评价维度,并构建了相应的多模态评价共同体运作机制。

在评价维度上,进行了具体的操作化定义。语言维度聚焦译语产出质量,涵盖信息忠实度、语言准确度、口语流利度及语用适切性。非言语维度关注副语言信息与体态语,包括语音特质、目光接触、面部表情与身体姿态等体现专业性与沟通效能的要素。认知维度评估学生的心智策略与负荷管理能力,通过反应延时、信息压缩策略、失误修复能力等指标予以考察。情感维度则着重关注学生的学习心理状态,如情境焦虑水平与自我效能感。

在运作机制上,构建了由机器自动评价、学生自我评价、同伴评价及教师评价组成的协同系统。机器负责对可量化指标进行高速、客观的分析;学生自我评价引导其对认知策略与情感状态进行反思;同伴评价提供非言语表现及协作精神的观察视角;教师则作为评价的最终整合者,对高阶能力进行质性评判与个性化指导。这种分工协作机制有效保障了评价的全面性与深度。

#### 3.3 技术实现路径与实施流程

该体系的顺畅运行依赖于一套精心设计的数智化工具链与标准化的实施流程。从技术路径来看,体系采用模块化架构:数据采集模块通过集成音视频录制与日

志追踪的智能平台, 同步捕获多模态数据; 分析处理模块运用语音识别与自然语言处理引擎分析语言指标, 通过计算机视觉模型解析非言语行为; 可视化反馈模块将分析结果整合为多模态学习分析报告, 向不同主体呈现。

实施流程遵循严格的标准化程序。首先, 学生在智能平台上完成同传任务, 系统自动采集音视频流及操作日志数据。随后, 数据进入自动化分析阶段, 生成包含各维度指标的初步评价报告。接着, 报告被推送至学生端用于自我评价与同伴互评。最后, 教师端接收整合所有数据的综合报告, 据此进行专业评判与教学干预。这一闭环流程确保了从数据采集到教学决策的有效转化, 实现了从静态分数到动态能力画像的范式转移。

#### 4 实证研究: 在“同声传译”课程中的实施与效果验证

##### 4.1 研究设计

为验证所构建的数智化多模态评价体系的有效性 with 实用性, 本研究在枣庄学院外国语学院英语本科三年级(2023级)开展了一项为期一个学期的教学实证研究。研究采用基于设计的研究范式与混合研究方法, 旨在通过量化数据与质性资料相结合的方式, 全面探究该评价体系对教学过程与学习成果的实际影响。

研究选取口译课程一个自然教学班共计35名学生作为研究对象, 他们已先修过视听说及笔译类课程, 具备了较高的听说和双语转换能力。研究情境即为该课程的正常教学流程, 研究覆盖了从信息密集型演讲到技术性讲座等四个核心教学单元。

实证研究的实施紧密围绕所构建的评价体系展开。在每个教学单元中, 学生需在集成了音视频录制与日志追踪功能的智能教学平台上完成特定的同传任务。任务结束后, 系统自动生成包含语言流利度、准确度、反应延时及非言语行为初步分析的多模态学习分析报告, 并即时推送至学生端。学生需首先根据该报告进行自我评价, 反思其在认知策略运用与任务过程中的情感体验, 并填写结构化反思表单。随后, 进入同伴评价环节, 学生以小组形式通过平台互相审阅任务视频, 重点关注并评价彼此的非言语表现(如眼神接触、专业姿态)及信

息处理的有效性, 并提供质性反馈意见。最后, 教师端将接收到整合了机器数据、学生自评与同伴评价的综合报告。教师以此为基础, 结合自身的专业观察, 对学生的信息忠实度、语用適切性等高阶能力进行最终评判, 并形成个性化的改进建议。整个评价流程贯穿于每个教学单元, 实现了对学习过程的持续追踪与形成性反馈。

##### 4.2 数据收集与分析

为全面评估数智化多模态评价体系的实施效果, 本研究采用混合研究方法, 从量化与质性两个层面收集数据, 并通过三角互证确保研究信度与效度。

###### (1) 数据来源与类型

数据主要来源于以下四个方面: 1) 平台自动生成数据: 智能教学平台在每次任务后自动生成的多模态学习分析报告, 涵盖语言维度(如流利度、准确度、信息忠实度得分)、非言语维度(如目光接触频率、姿态稳定度、面部表情丰富度)、认知维度(如平均反应延时、信息压缩比率、修正次数)的量化指标。2) 学生自评与互评数据: 包括学生在每个单元完成后提交的结构化反思表单(含认知策略描述与情感状态自评)以及同伴互评的质性反馈与评分。3) 教师评价数据: 教师基于综合报告给出的最终评分与质性评语, 重点关注语言高阶能力(如语用適切性、跨文化意识)和整体表现。4) 访谈与问卷数据: 研究中期和末期分别对10名学生进行半结构化访谈, 了解其对评价体系的感知、体验与建议; 并在学期末对全班学生发放“评价体系接受度与效果调查问卷”(采用李克特五级量表, Cronbach's  $\alpha=0.87$ ), 收集态度与自我感知数据。

###### (2) 数据分析方法

量化分析: 使用 SPSS 26.0 对平台数据、评价分数与问卷数据进行统计分析, 包括描述性统计、配对样本 t 检验(比较不同阶段能力指标的变化)、皮尔逊相关性分析(检验不同评价主体得分间的一致性)。

质性分析: 对反思文本、同伴反馈、教师评语及访谈转录文本进行主题分析, 提取关于学生认知策略运用、情感体验变化、对多模态反馈的感知等关键主题, 以解释量化结果背后的深层原因。

表 1 数据收集类型与方法一览表

| 数据类型    | 具体内容                 | 收集工具/方式     | 分析方法       |
|---------|----------------------|-------------|------------|
| 多模态表现数据 | 语音流利度、准确度、反应延时、目光接触等 | 智能教学平台自动采集  | 描述性统计、t 检验 |
| 学生反思与自评 | 认知策略描述、情感状态自评        | 结构化反思表单     | 主题分析       |
| 同伴互评    | 非言语表现评价、协作能力反馈       | 平台互评模块      | 内容分析、评分统计  |
| 教师评价    | 高阶能力评分、质性评语          | 教师评分界面      | 描述性统计、主题分析 |
| 问卷与访谈   | 体系接受度、自我效能感、满意度      | 问卷调查、半结构化访谈 | 量表分析、主题编码  |

表2 四个教学单元关键指标均值变化 (N=35)

| 评价维度 | 关键指标          | 单元 1  | 单元 2  | 单元 3  | 单元 4  |
|------|---------------|-------|-------|-------|-------|
| 语言   | 信息忠实度 (满分 10) | 6.24  | 6.78  | 7.30  | 7.85  |
| 语言   | 口语流利度 (WPM)   | 112.3 | 118.5 | 123.9 | 128.7 |
| 非言语  | 专业姿态保持 (秒/分)  | 38.2  | 42.7  | 46.1  | 49.6  |
| 认知   | 平均反应延时 (毫秒)   | 1856  | 1689  | 1543  | 1423  |
| 情感   | 自我效能感 (5 级量表) | 3.12  | 3.45  | 3.68  | 3.89  |

#### 4.3 研究结果

##### (1) 多维度能力发展呈显著提升

通过对四个教学单元的平台数据进行纵向追踪分析发现,学生在语言、非言语、认知三个维度的关键指标上均呈现明显且持续的进步(见表2)。语言维度的信息忠实度(单元1vs.单元4:  $t=4.32, p<0.01$ )与口语流利度(单元1vs.单元4:  $t=3.87, p<0.01$ )提升明显;非言语维度中,专业姿态保持时间显著增长(单元1vs.单元4:  $t=5.12, p<0.001$ );认知维度的平均反应延时显著缩短(单元1vs.单元4:  $t=4.89, p<0.001$ ),表明信息处理效率随着训练逐步提升。情感维度自评数据显示,学生的情境焦虑水平从单元2后呈下降趋势,而自我效能感得分则随任务推进稳步上升。

##### (2) 多主体评价协同有效,信度提升

相关性分析显示,机器自动评价、学生自评、同伴评价与教师评价在语言维度和非言语维度上的得分具有中度至高度正相关( $r=0.62\sim 0.81, p<0.01$ ),表明多主体间具有较好的一致性,体系信度良好。质性分析进一步揭示,学生普遍认为同伴互评“提供了新的观察视角”,教师评语“更具针对性”;教师反馈则指出,机器提供的客观数据“减少了评分主观性,使反馈更具依据”。

##### (3) 过程性反馈促进元认知与策略运用

对反思文本的主题分析发现,随着评价流程的循环进行,学生描述认知策略的深度与频率明显增加。早期反思多集中于“听不懂”“紧张”等表层描述,后期则更多出现“我尝试了信息压缩”“通过预判缓解了延时”等策略性表述。访谈中多位学生提到:“看到自己每个单元的数据变化,更清楚该在哪方面用力”。这表明,持续的多模态过程性反馈有效提升了学生的元认知意识与策略调节能力。

##### (4) 人机协同提升评价效能与教学针对性

问卷结果显示,85.7%的学生认为该评价体系“比

传统方式更全面、公平”;教师反馈显示,评价时间成本较传统方式减少约40%,且能更早识别学生的个体瓶颈。教师得以将更多精力投入到设计个性化干预与高阶能力指导,体现了“人机协同”在提升评价效能与教学针对性方面的双重优势。

#### 5 讨论与推广路径

本研究构建并验证的数智化多模态评价体系,为破解外语教学评价,特别是高端听说类技能课程的评价困境提供了系统性方案。实证结果表明,该评价体系有效回应了引言所提出的“三重困境”。首先,通过确立“语言-非言语-认知-情感”四个维度,并借助技术手段实现其可观测与可测量,体系突破了传统评价的维度窄化局限,使学生综合素养的全面评估成为可能。其次,将评价嵌入从数据采集到多轮反馈的学习全过程,实现了从重结果到过程与结果并重的范式转移,为学生能力的动态发展提供了精准“导航”。最后,人机协同的评价机制不仅显著提升了评价效能,将教师从繁琐的重复性劳动中解放出来,更通过机器的客观数据与人的质性判断相结合,增强了评价的科学性与公正性。

本研究实现了“多模态”、“数智化”与“人机协同”三者的有机统一,并将三者系统整合为一个连贯的评价生态。其中,“人机协同”不仅是技术层面的结合,更是评价权重的重新分配与评价关系的重构。它要求教师从唯一的评判者转变为评价活动的设计者、机器数据的解读者与学生发展的引导者,这对教师的评价素养提出了更高要求,也构成了实施过程中的主要挑战。

本研究亦存在若干局限性。因实证周期相对有限,该评价体系对学生专业能力发展的长期影响尚需追踪。此外,体系的有效运行在一定程度上依赖于技术平台的稳定性与师生对新模式的适应性,这在技术保障不足或师生信息素养参差不齐的情境下可能制约其推广效果。这些挑战同时也指明了未来研究的方向:一方面

需进一步探索如何优化技术算法, 提高其对复杂语言现象和跨文化交际行为的解读能力; 另一方面, 应着手开发配套的师生工作坊与指导资源, 降低使用门槛, 促进评价理念的真正内化与实践转化。

## 6 结语

本研究以数智化时代背景下的外语教学评价改革为切入点, 聚焦同声传译课程, 系统构建并实证检验了一套多模态评价体系。研究结论如下: 第一, 融合多模态学习理论、现代语言能力模型与数智化技术的理论框架, 为语言评价研究提供了新的理论视角。第二, 所设计的涵盖四个维度、四方主体的评价体系, 在实践中被证明能够有效促进教学评一体化, 提升学生的学习成效与自我效能感。第三, 人机协同的路径为解决评价的效率与信度难题提供了可行方案。

基于本研究, 作者建议未来可采取“由点及面”的策略, 首先在具备条件的同类高端技能课程中应用, 逐步积累经验后, 再结合不同课程特点进行适应性调整, 推广至其他语言技能课程。同时, 还应同步推进技术支持、师资培训与课程改革的“三位一体”支持策略, 为体系的落地实施提供坚实保障。

## 参考文献

- [1] Housen, A., & Kuiken, F. (2025). The architecture of automatic scoring systems for non-native English spontaneous speech: A systematic literature review. *Journal of Intelligent Systems and Applications*, 7(1), 1-22.
- [2] 文旭. (2025). 未来外语教育中的机器人教师: 机遇与挑战. *中国外语*, 22(1), 1,10-16.
- [3] Collard, C., & Defrancq, B. (2019). Disfluencies in simultaneous interpreting: A corpus-based study with special reference to sex. *Interpreting*, 21(2), 223-255.
- [4] Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Routledge.
- [5] Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- [6] Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, 2(1), 29-57.
- [7] Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- [8] Larsen-Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal*, 103(SI), 61-79.
- [9] Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3-14.
- [10] 金霞. (2012). 工作记忆容量限制对二语学习者口语产出的影响. *外语教学与研究* (外国语文双月刊), 44(4), 523-535.

**版权声明:** ©2026 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**