

基于CSE的智能评估在中职高考英语写作教学中的实证研究

梁彩云¹, 许妍²

¹上海市群星职业技术学校英语组, 上海

²上海市浦东外事服务学校师训处, 上海

收稿日期: 2025年11月26日; 录用日期: 2026年1月21日; 发布日期: 2026年1月28日

摘要

本研究旨在解决传统中职英语写作教学中教师批改负担重、反馈不及时、缺乏个性化指导的问题。为此, 研究者构建并实践了一种融合《中国英语能力等级量表》(CSE) 与生成式人工智能的“四阶段两动态介入”智能动态评估模式。将该模式应用于一个为期一学期的中本贯通二年级班级的高考英语写作教学实验中, 采用单一组前后测设计, 通过写作任务、问卷调查等方式收集数据, 运用配对样本t检验等方法进行分析。研究结果表明, 该教学模式显著提升了学生的整体写作成绩以及在内容、语言、结构等维度的单项得分, 尤其对学业后进生的促进作用更为明显。研究表明, 该模式为实现写作教学的智能评价与“以评促学”提供了有效路径。

关键词

中国英语能力等级量表(CSE), 智能评估, 中职高考英语写作

An Empirical Study on Intelligent Assessment Based on CSE in the English Writing Teaching of Vocational High School Students for College Entrance Examination

Caiyun Liang¹, Yan Xu²

¹English Department, Qunxing Vocational Technical School, Shanghai

²The Teaching Training Department of Shanghai Pudong Foreign Affairs Service School, Shanghai

Abstract

This study aims to address the problems of heavy teacher workload in correcting, delayed feedback, and lack of personalized guidance in traditional vocational high school English writing teaching. To this end, the researchers have constructed and implemented a “four-stage two-dynamic-intervention” intelligent dynamic assessment model that integrates the “China’s Standards of English Language Ability” (CSE) and generative artificial intelligence. This model was applied to a one-semester teaching experiment of College Entrance Examination English writing for a second-year class in a secondary vocational school. A single-group pretest-posttest design was adopted, and data were collected through writing tasks and questionnaires. Paired sample t-tests and other methods were used for analysis. The research results show that this teaching model significantly improved the overall writing scores of students and their scores in content, language, and structure. The improvement was particularly significant for students with lower academic performance. Research shows that this model provides an effective path for achieving intelligent evaluation in writing teaching and promoting learning through evaluation.

Keywords

China’s Standards of English Language Ability (CSE), Intelligent Assessment, Vocational High School College Entrance Examination English Writing

Copyright © 2026 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. 引言

我国现代职业教育体系不断完善, 中职教育作为职业教育基石, 其人才培养质量关乎国家技能人才队伍建设与学生发展。在这一背景下, “文化素质 + 职业技能” 的职教高考制度成为中职生升学深造的主渠道, 而英语作为重要的公共基础课程, 其考核成绩对考生的升学结果具有决定性影响。基于此, 英语写作能力的重要性、中职生写作能力的现状与其培养的紧迫性, 构成了本研究展开的核心背景。

英语写作在“职教高考”中地位关键, 既是考查语言基础的标尺, 也是评估逻辑思维等能力的载体, 优质答卷能提升考生竞争力; 同时, 良好的英语书面表达能力也是学生后续学习专业知识和职业发展的必备素养。

传统英语写作教学模式较难关注个体需求, 其问题主要表现为:

1) 反馈评价粗、时效性差: 由于班级规模的限制, 教师难以对每位学生的作文进行精细、及时的批改[1] [2]。

2) 作文集体讲评, 缺乏个性化指导: 在传统模式下教师难以做到对每个学生进行面批。由于缺乏个性化的、基于标准的指导(如 CSE), 学生只能依赖模糊的集体讲评或模仿范文。集体讲评虽然能在大方面如内容上确保作文不偏题, 但无法贴合学生自己的文本在词汇和语法基础上修改完善。而参考借鉴老师所给范文则导致作文同质, 缺乏个性。从而难以实现基于自身水平的有效提升和个性表达[3]。

而智能批改系统如豆包、文心一言、Kimi、DeepSeek 以及 ChatGPT 等都需要有规范统一详细的作文

批改提示词。而提示词输入不同的评价标准、考量角度和分值占比, 就会生成不同的评价分数。智能批改模型本身并不具备评价能力, 其评判逻辑完全由使用者通过提示词来规定。若提示词中的评价标准模糊、片面, 生成式智能的输出便会不稳定。CSE 的书面表达能力量表及其配套策略量表, 恰好能为智能批改提供一套统一的标准。

2. 文献综述

《中国英语能力等级量表(CSE)》是第一个面向我国学习者的英语能力标准。目前关于 CSE 的研究主要侧重理论构建与效度验证、在教学中的实际应用和在测评与考试改革中的应用研究。

CSE 的构建以维果斯基的社会文化理论为基础, 强调“最近发展区”(ZPD)在学习中的核心作用[4]。学者们充分验证了其信度和效度。如, 邓华、邓杰(2022)[5]证实 CSE 写作策略量表具有良好的信度和构念效度。

CSE 的应用研究覆盖小学至大学各个学段, 主要探索了量表在写作策略指导和形成性评价构建等方面。大量研究证实其对提升学生写作能力与自我评价具有积极作用。然而现有研究缺乏将 CSE 量表与生成式技术、动态评估系统等智能工具深度融合的探索。

动态评估理论在英语写作教学中的应用研究的核心是通过介入性活动来诊断和发展学生的写作潜能。动态评估理论也是以维果斯基的“最近发展区”理论为基础, 强调对个体能力的评估不应局限于其独立解决问题的“实际发展水平”, 更应关注在成人或与同伴协作下所能达到的“潜在发展水平”[6]。因此, 动态评价的本质是一种通过介入来评估和发展个体潜能的过程[7]。

然而, 动态评价的实施也需要教师投入大量时间和精力进行教学设计, 对每个学生面批作文并提供个性化反馈。这在大班教学中很难实现。

CSE 为动态评价提供了科学、统一的操作框架。CSE 量表中“能做”的限定描述语可直接转化为学生自评与同伴互评的具体指标, 使评价更加有据可循。笔者将 CSE 中写作的核心维度如内容充实度、逻辑连贯性、语言准确性及其相应等级描述语融入提示词, 确保智能评价的一致性和稳定性。如: 对“逻辑连贯性”的评价, 可以依据 CSE 中从“能使用简单的连接词(三级)(依据中国英语能力等级量表 5-16 和 5-14)[8]”到“能综合运用多种衔接手段, 使全文连贯流畅(四级)”的渐进式描述来设定提示词, 依此给出具体的分数, 并给以详细的范例, 使智能评分更精确。同时, 基于 CSE 制定了自评和互评表辅助学生进行自评和互评, 将教师从繁重的细节批改中解放出来, 使其更专注共性问题的精准指导和对薄弱学生的个性化介入。这有效缓解了动态评估在大班教学中难实施、教师负担重的困难。

3. 基于 CSE 的智能动态评估模型构建

教师负担重 基于上述理论与教学实践, 本研究构建了“四阶段两动态介入”的智能动态评估模型。见图 1。

本教学流程以《中国英语能力等级量表》(CSE)为根本依据, 融入生成式技术与动态评估理念, 划分为写前、初稿、二稿和终稿四个阶段。具体流程如下。

3.1. 写前准备阶段

教师借助生成式技术, 围绕写作主题, 针对性搭建适配的写作词汇库, 为学生后续写作筑牢基础。

3.2. 初稿阶段

学生在课堂内完成初稿创作。之后教师对学生作品拍照, 在豆包中输入详细的评分标准, 分别对学生作品的语篇内容、语法结构、语言表达、组织结构进行评价。之后将智能评估给出的详细的纠错和提

升建议逐个发给学生,并要求学生结合 CSE 制定的自评量表进行自评,使学生清晰认知到自身写作的短板。见表 1:

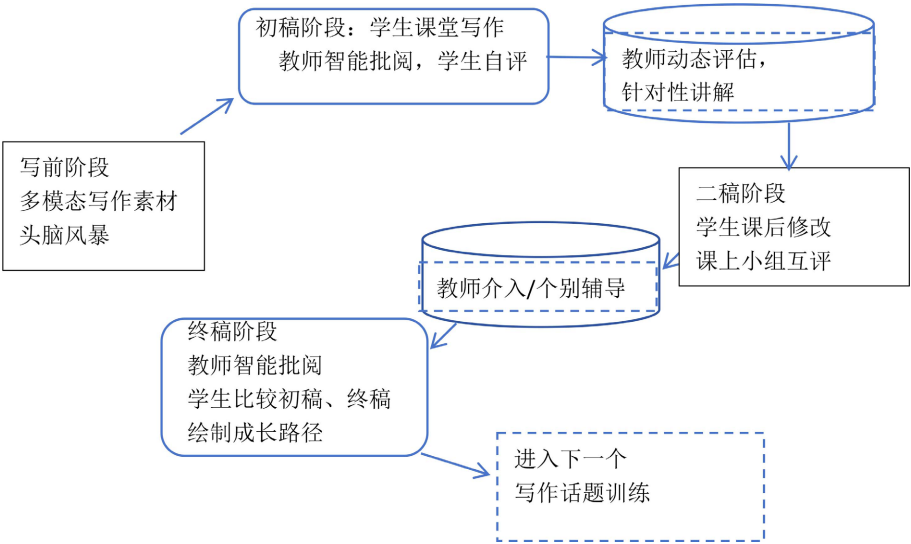


Figure 1. Intelligent dynamic evaluation model based on CSE diagram
图 1. 基于 CSE 的智能动态评估模型图

Table 1. The self-assessment scale formulated based on CSE—content dimension
表 1. 基于 CSE 制定的自评量表——内容维度

一、内容维度(满分 6 分)		
分值区间	评分标准	CSE 量表对应依据(表 5-14、5-16)
5~6 分	1) 完全贴合作写要求,所有核心内容点均覆盖,无遗漏;2) 描写细腻,有具体细节支撑(如实例、场景、心理等);3) 语言流畅自然,逻辑清晰;4) 每句或关键句含高中阶段高级词汇(如 delicate, enthusiastic, accumulate 等),词汇运用准确恰当。	对应 CSE 5~6 级:能围绕主题完整传递信息,内容充实且细节丰富;词汇运用精准,能根据语境选用较复杂词汇,语言表达流畅自然。
3~4 分	1) 基本贴合作写要求,但缺失 1 个核心内容点;2) 描写较笼统,细节支撑不足;3) 语言比较流畅,无明显逻辑断层;4) 含部分高中阶段高级词汇(如 significant, approach, challenge 等),运用基本准确。	对应 CSE 4 级:能围绕主题传递主要信息,内容较完整但细节不够丰富;词汇运用基本准确,能使用一定量复杂词汇,语言表达较流畅。
1~2 分	1) 偏离写作要求,核心内容点缺失 2 个及以上;2) 语言欠流畅,内容混乱或重复,表达模糊;3) 层次不清晰,句间衔接生硬;4) 基本无高中阶段高级词汇,多使用基础词汇且部分运用不当。	对应 CSE 2~3 级:内容偏离主题或关键信息缺失;词汇以基础词汇为主,运用存在不当;语言表达欠流畅,逻辑连贯性较弱。
0 分	内容与写作要求完全无关,或仅为零散句子堆砌,无法表达完整含义。	低于 CSE 2 级:无法围绕主题传递有效信息,语言表达碎片。

语法维度和结构维度均依据 CSE 量表进行细化, 语法维度细分为 4 个档次, 共 6 分, 占分值 40%。组织结构共 3 分, 占总分值 20%。此处不再赘述。

3.3. 动态评估介入, 针对性讲评

此环节作为动态评估的关键节点。教师需基于前期智能评价与学生自评的数据, 聚焦学生在词汇运用丰富性、句法结构多样性及组织结构合理性方面存在的共性问题, 通过针对性讲解提供干预支持。如针对句式简单的共性问题, 进行简单句变从句或分词的练习; 针对语篇衔接不通顺的问题, 则进行连接词的巩固训练。同时, 选取学生作品中的优秀片段或完整佳作进行展示与评析, 在指出问题、提供帮助的同时, 发挥正向引导作用, 为学生后续写作优化提供可参考的“支架”范例。

3.4. 二稿打磨阶段

课后, 学生依据自己的初稿和智能批阅的详细反馈对文稿内容、语言和结构进行优化完善。完成修改后学生将教师发布的详细指令输入智能评估, 获得第二次智能反馈意见, 并将反馈小分、总分和建议记录在作文纸上。随后, 课堂上以小组为单位, 依据 CSE 互评表对其他组员的文稿进行互评, 并记录互评建议和分数。

教师根据互评结果实施分层指导策略: 对于已有效解决写作问题的学生, 引导其完成终稿; 对仍面临写作瓶颈的学生, 则需提供针对性指导, 以帮助其明确后续提升路径。

3.5. 终稿——总结阶段

学生完成终稿并提交后, 教师再次利用智能批阅, 使用同样的评价标准, 对学生作品进行评价, 最后将结果反馈给同学。最后要求同学记录初稿和终稿每个评价版块的分值和总分, 分析此次写作收获最大之处。

4. 研究设计

4.1. 研究对象

本研究选取本校中本贯通培养模式下高中二年级的一个班级作为研究对象, 该班级共 30 名学生。由于本年级仅设置 1 个中本贯通班级, 与其他班级在学生能力水平上存在显著差异, 不具备设置实验班与对照班开展对比实验的条件, 难以进行横向比较, 因此以该班级的实验前和实验后做比较, 开展了为期一学期的教学实验研究。

4.2. 研究工具

研究工具包含写作任务与问卷调查。写作任务要求学生完成四个不同话题的书面描写文章的撰写, 话题分别为“The Greatest Change in My Growth”“An Unforgettable Experience”“The Person I Show Gratitude”“The Early Bird Catches the Worm”。其难度依据 CSE 量表四级难度的要求设置; 问卷调查则用于调查学生对智能动态评估工具、学生互评和教师评价的接受度以及在写作具体方面的提升度。

4.3. 研究过程: 开展基于 CSE 的智能动态评估写作教学

4.3.1. 第一环节: 前测(第 1~2 周)

本阶段旨在通过初始能力诊断建立教学实践的基线参照。为清晰把握学生的实际写作水平, 所有学生需完成一项基于 CSE 四级书面描述能力的写作任务(前测)。研究开展前, 已通过书面形式向学生及家长详细说明研究目的、流程与数据使用原则, 并征得其知情同意。随后, 利用智能批阅系统对前测作文

进行自动化评分与分析,生成在语言准确性、词汇复杂度及篇章连贯性等方面的量化数据。在此基础上,引导学生参照基于 CSE 书面描述能力量表改编的自评量表,对作文的内容充实度与结构清晰度等进行反思性评价,以形成对自身写作能力的初步认知(见附录一)。

4.3.2. 基于 CSE 的智能动态评估模型的四个话题的写作训练阶段(第 3~14 周)

本阶段共包含四个完整的写作教学循环。每个循环聚焦一个具体的书面描述话题,并遵循统一的四阶段两动态介入的流程。

1) 写前阶段:教师利用智能工具为学生提供多模态语言素材,生成需要的词汇库。如以 The Greatest Change in My Growth 作文题目为例,豆包生成如下多场景素材库。素材库如下所示:

1) 学习能力提升场景

核心变化:从“不爱学习”到“喜欢学习”

素材细节: Before, I didn't like studying. I tried to read books and do exercises. Now, I like learning new things. My grades are better.

仿写示例: The greatest change in my growth is loving study. Before, I thought study was boring. My teacher told me to try reading simple stories. I did it and found it interesting. Now, I like to learn English and math. I feel happy when I get good grades.

2) 性格与社交成长场景

核心变化:从“害羞”到“开朗”

素材细节: I was shy before. I joined the school club. I talked to classmates and played games with them. Now, I have many friends and I am happy.

仿写示例: The greatest change in my growth is being outgoing. I used to be afraid to talk to others. Then I joined the art club. I drew pictures with my classmates and shared ideas. Now, I can talk to anyone easily.

3) 自我管理成长场景

核心变化:从“拖延”到“自律”

素材细节: I used to delay homework. I made a plan to do homework first every day. I followed the plan and stopped playing games too much. Now, I finish work on time and feel good.

仿写示例: The greatest change in my growth is being disciplined. Before, I always played games and forgot homework. My mom helped me make a daily plan. I did homework first and then played. Now, I stick to the plan. I finish all tasks on time and feel proud of myself.

2) 初稿阶段:学生在课堂内完成初稿写作,教师根据智能批阅反馈的普遍问题进行动态介入。以下是一位程度稍弱学生的作文反馈图,从图中可以看出豆包智能评估给出的三大版块具体的问题以及建议。见图 2:

(一) 内容方面

优点：提及了成长前后的变化(之前依赖祖母，之后学会照顾家人)以及转变的原因(祖母受伤住院，自己开始独立解决问题)。

不足：没有清晰阐述为变化所付出的努力，属于内容不全面，缺失一个方面的内容。

(二) 语法方面

错误举例：

“When I was Grade 4.” 应改为 “When I was in Grade 4.”，缺少介词 “in”，属于介词误用，扣 1 分。

“My grandmother took an car accident” 应改为 “My grandmother had a car accident”，“take” 使用错误，且 “car” 是以辅音音素开头，不定冠词用 “a”，属于动词和冠词误用，扣 2 分。

“Before she hurt” 应改为 “Before she got hurt”，“hurt” 作 “受伤” 讲时，常用 “get hurt” 结构，属于动词用法错误，扣 1 分。

“I always very tired” 应改为 “I was always very tired”，缺少 be 动词 “was”，属于句子结构错误，扣 1 分。

“I knowed how tired my grandma was” 应改为 “I knew how tired my grandma was”，“know” 的过去式是 “knew”，属于动词过去式错误，扣 1 分。

“After she come back home” 应改为 “After she came back home”，“come” 的过去式是 “came”，属于动词过去式错误，扣 1 分。

语法错误较多，超出 4 个，所以语法得分在 2 - 4 分区间，最终给 2 分。

(三) 结构方面

不足：文章整体为一段，句子之间连接比较生硬，只有 “Then” “So” 等少量过渡词，所以结构得分在 1 - 2 分区间，最终给 1 分。

Figure 2. Intelligent evaluation score feedback graph
图 2. 智能评价评分反馈图

教师进一步汇总所有同学的分数，并分析核心问题进行动态介入。具体见表 2：

Table 2. The scoring rates of each dimension of the students’ initial draft essays and the analysis of the problems
表 2. 学生初稿作文各维度得分率以及问题分析

版块	均分	总分	得分率	核心问题分析	动态介入教学提升方案
内容	3.63	6	60.5%	1) 内容不够具体，缺乏细节描写(如仅提及“有变化”，未展开具体事件)；2) 主题关联度不足，部分内容偏离“成长变化”核心。	1) 结合之前的仿写素材，引导学生补充场景细节(如“学习能力提升”需写清“具体做了什么”)；2) 课前明确作文主题关键词，要求每段内容紧扣“变化”展开。
语言表达	2.36	6	39.3%	1) 基础词汇运用单一，句式重复(多为简单句，缺乏基础复合句)；2) 语法错误较多(时态、主谓一致、介词使用不当)。	1) 强化“词汇库”背诵与运用，要求每篇作文至少使用 5 个核心词汇；2) 句式训练，将简单句变为从句或分词。(整理学生作文中的简单句进行练习)。
语篇衔接	1.48	3	49.3%	1) 段落之间缺乏过渡词(如并列、转折、递进、因果、总结等连接词)；2) 句子衔接生硬，逻辑不连贯(如突然切换场景，无铺垫)。	1) 整理“基础衔接词清单”，要求学生每篇作文至少使用 3 个；2) 用仿写素材中的“变化过程”段落为范例，讲解“先写过去→再写转变→最后写结果”的衔接逻辑。

由表 2 看出，语言表达版块问题最大(得分率仅 39.3%)，其次是语篇衔接(49.3%)，内容版块表现相对较好(60.5%)。针对学生作文在语言表达、语篇衔接及内容呈现方面存在的共性问题，教师设计并开展了

三节专项提升课。具体教学包括：从句运用与句式变换专项训练、语篇衔接词的精准使用指导，以及同一主题不同写作场景下的框架搭建方法实操练习。通过教师的动态介入和辅导，学生能更好更清晰地进行二稿的改进。

3) 二稿同伴互评与教师个别介入阶段

学生基于智能反馈意见对初稿进行修改，形成二稿。随后，在课堂中以小组为单位，学生开展互评活动(见附录二)，深化对评价标准的理解与实际应用。在此过程中，教师进一步对部分学习薄弱的学生进行个别化辅导，有针对性地引导其完善文稿，最终促成终稿的完成。

研究表明，在先前无人工智能辅助写作后，使用人工智能工具改写文章会引发更广泛的大脑网络互动[9]。而同伴互评与教师面批，帮助学生从“修改”迈向“优化”，尤其为写作薄弱学生提供关键支持。例如互评中有小组给出评语：“文中提到‘那一刻我长大了’，但前后的心理对比可以更强烈些，描写地再细腻一些。”

教师快速汇总互评分数，对分数偏低的学生进行一对一的面批指导，引导其发现问题。如一名学生选择“第一次独自离家住宿”作为“巨变”主题。但仅描写了：“I missed my home so much that I cried.”教师面批时，没有直接给出优美词句，而是引导学生回忆具体的感官片段，不断补充细节，增加感官和动作描写以充实内容。学生体悟到要从不同角度增加内容，并使用恰当的衔接词过度，以达到写作能力的提升。

4) 终稿——总结反思阶段

学生结合互评意见形成终稿。教师再次根据同一标准利用智能评价给出终稿分数。最后教师要求学生系统记录并对比初稿与终稿在各板块的得分及总分，并书面分析此次写作过程中收获最大之处。这一反思环节旨在将感性的写作体验转化为可量化的成长数据，帮助学生清晰感知自身进步，建立写作自信。

4.3.3. 后测与效果评估(第 15~16 周)

实验后写作能力评估：以第四个书面描写文的学生初稿成绩作为实验后测成绩，以评估其经过一个学期的教学实验后的学生的写作能力是否得到提升。初稿是学生当堂完成，没有教师的辅导和智能工具的帮助，完全考察学生实际的写作水平。且文章的难易度符合 CSE 中四级考察的难度，学生也没有写过同一话题的作文，所以数据有一定真实性。

在教学实验结束后，对所教授班级学生进行实施问卷调查，了解学生对智能动态评估工具的接受度和学生写作具体方面的提升程度。

5. 结果与分析

5.1. 前测和后测的配对样本 t 检验

Table 3. Paired test of pretest and post-test
表 3. 前测与后测的配对检验

前测与后测作文 的配对检验	前测均值 ± 标准差	后测均值 ± 标准差	差值	t 值	p 值	Cohen's d 值
作文总分 15 分	8.60 ± 1.77	12.03 ± 2.09	-3.43	-9.866	0.000**	1.801
内容总分 6 分	3.70 ± 0.79	5.03 ± 0.89	-1.33	-7.616	0.000**	1.390
语言表达 6 分	2.90 ± 0.96	4.17 ± 0.83	-1.27	-7.077	0.000**	1.292
语篇结构 3 分	1.08 ± 0.61	2.81 ± 0.46	-1.03	-7.878	0.000**	1.438

* $p < 0.05$; ** $p < 0.01$ 。

从表 3 可以看出, 经过一个学期的教学实践, 学生的英语写作能力在整体及各个分项维度上均取得了统计学上的显著提升。具体而言:

整体写作水平: 作文总分呈现出极显著的进步($t = -9.866, p < 0.001$)。后测平均分(12.03)相较于前测平均分(8.60)提升了 3.43 分, 增幅达 39.9%, 表明学生的综合写作产出能力有了显著提高。内容构思能力: 内容维度得分差异显著($t = -7.616, p < 0.001$)。后测得分(5.03)比前测得分(3.70)提高了 1.33 分, 这反映出学生在切题性、内容充实度与观点阐述方面取得了明显进步。语言运用能力: 语言表达得分亦呈现极显著差异($t = -7.077, p < 0.001$)。均值从 2.90 提升至 4.17, 增长 1.27 分, 说明学生在词汇选择、句式复杂度与语法准确性等方面的操控能力得到有效增强。语篇组织能力: 语篇结构维度的提升幅度最为突出, 后测得分(2.81)几乎是前测得分(1.08)的 2.6 倍, 均值差为-1.73 ($t = -7.878, p < 0.001$)。这表明教学干预尤其有效地帮助学生改善了文章的逻辑连贯性与结构层次感。Cohen's d 值大于 0.8, 也说明干预效果较大。

5.2. 学生进步成长曲线分析

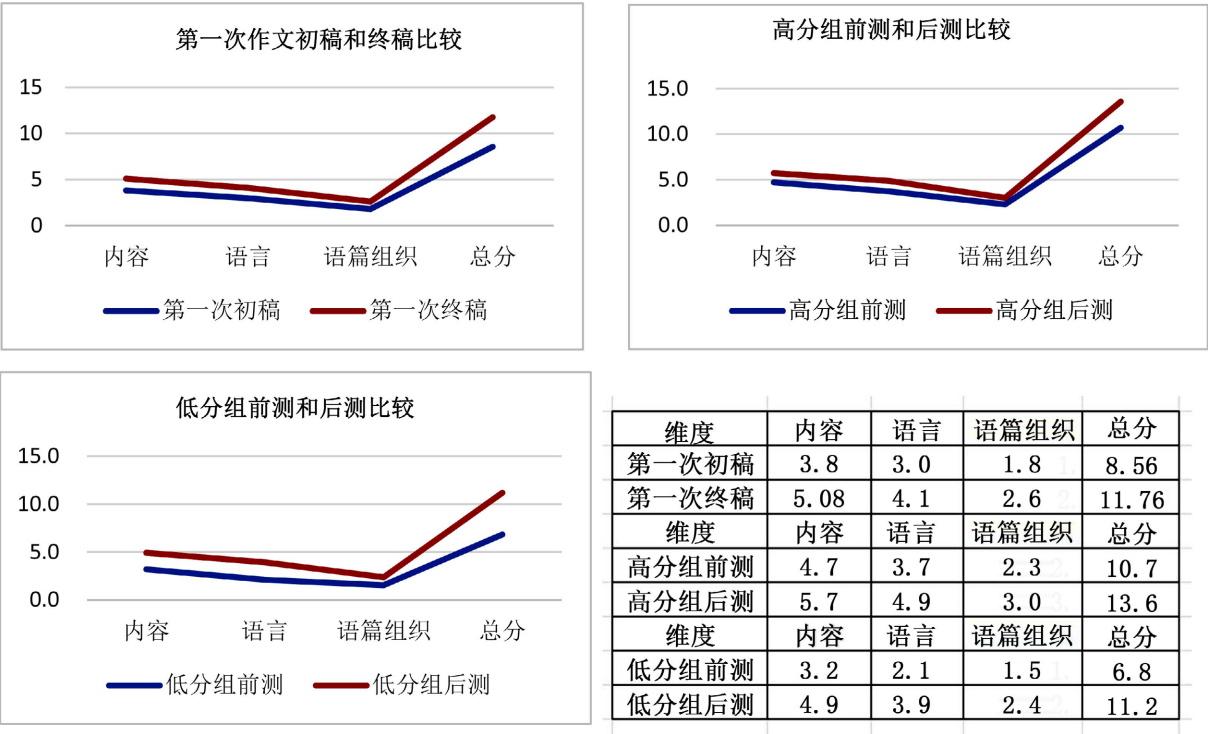


Figure 3. Comparative analysis of students' first drafts and final versions of their first writings
图 3. 学生第一次写作初稿和终稿对比分析

根据图 3 可以看出, 实验后, 全体学生的作文终稿总分(11.76)相较于初稿(8.56)有显著提升, 增幅达 37.4%。高分组与低分组学生的初稿成绩存在显著差距(总分差: $10.7 - 6.8 = 3.9$ 分), 尤其在语言(差值 1.6 分)和语篇组织(差值 0.8 分)方面差距更为突出。然而, 经过教学干预, 两组学生的终稿成绩均获得实质性提升。低分组进步幅度尤为显著, 其总分增幅高达 64.7% ($6.8 \rightarrow 11.2$), 且在内容(增幅 53.1%)和语言(增幅 85.7%)两个维度上的提升速率超过了高分组。

5.3. 调查问卷结果分析

为探究学生对作文智能批阅系统在英语写作学习中效用的主观感知与接受度, 本研究在实验结束后

对全班学生进行了问卷调查。问卷采用李克特五级量表(1= “非常不同意”, 2= “不同意”, 3= “一般”, 4= “同意”, 5= “非常同意”)进行数据收集, 共回收有效问卷 30 份。统计结果如下表 4 所示。

Table 4. Student satisfaction survey regarding the intelligent essay grading system
表 4. 学生对作文智能批阅系统的满意度调查

调查维度	具体描述	平均值 (M)	标准差 (SD)	同意及非常同意占比(%)
词汇丰富度提升	认为系统能帮助我将低级词汇替换为高级词汇, 提升词汇丰富度。	4.31	0.71	89%
语法准确性提升	认为系统能有效识别并纠正我的语法错误, 提高语法准确性。	3.51	0.52	78%
与教师评价比较	认为智能评估比传统教师批阅更有效或更能帮助我提高作文能力。	3.02	0.61	59%
与同伴评价比较	认为智能评估比同伴互评更有效或更能帮助我提高作文能力。	4.01	0.81	80%

由表 4 看出, 学生高度认可智能批阅。具体而言, 学生普遍感到智能批阅系统在纠正语法错误和提供高级词汇替代方面更即时、精准。在批阅方式的横向比较上, 大多数学生(80%)认为智能评估优于同伴互评。在将智能评估与教师批阅进行比较时, 59%的学生认为智能评估更有效, 41%的学生认为教师的辅导非常重要。

此外, 采用该教学方法后, 笔者对学生单篇作文从内容、语法和篇章结构等各方面详批并提供针对性建议所需的平均时间由原先的 11 分钟显著缩短至 3 分钟, 从而在教学反馈质量不受影响的前提下, 有效减轻了教师批改作业的时间负担。

6. 结语

本研究表明, 基于 CSE 标准的智能评估与教师动态介入相结合的写作教学模式, 对提升高中英语写作教学的质量成效显著。该模式通过人机协同, 构建了一个高效、及时的反馈闭环。智能批阅从多个维度给学生提供详细的反馈建议, 并能辅导学生在原有作文基础上对词汇、句式和内容进行提升, 有效激发了学生的学习自主性与写作兴趣; 而教师针对学生的共性问题 and 薄弱学生的个性问题进行动态介入, 高阶引导学生从内容切题性、逻辑结构与描写深度等宏观层面进行改进, 实现了对学生写作过程的个性化辅导。同伴互评能提供贴近最近发展区的反馈, 协助学生识别问题、调整策略, 实现写作能力的协同内化。

实践证明, 此种优势互补的教学范式, 能将教师从繁重的重复性批改中解放出来, 从而能更专注于启发思维、引导策略、把握教学核心。该模式优化了教学资源的配置, 整体提高了教学效率与效能, 因此具有重要的实践推广价值。

然而, 本研究亦存在一定局限性。首先, 当前智能评估技术尚无法识别并修正所有错误, 其手写体识别功能在字迹潦草或修改痕迹较多时, 会出现将错误书写误判为正确语句的情况。其次, 本研究采用单组前后测设计, 学生数量也较少, 未来研究需设置实验班与对照班更大规模地验证该模式的有效性。

参考文献

[1] 张艳红. 大学英语写作教学的动态评价体系建构[J]. 解放军外国语学院学报, 2010, 33(1): 46-50.

-
- [2] 李奕华. 基于动态评估理论的英语写作反馈方式比较研究[J]. 外语界, 2015(3): 59-67.
- [3] 张文霞, 王晓琳. 中国英语能力等级量表在大学英语写作课堂学生自评中的应用研究——动态评价理论视角[J]. 中国外语, 2022, 19(1): 71-78.
- [4] 刘芹, 王可人. 基于 CSE 的大学英语写作动态评估模型构建及应用研究[J]. 外语与外语教学, 2024(5): 102-113.
- [5] 邓华, 邓杰. 《中国英语能力等级量表》写作策略能力量表的构念效度验证[J]. 外语与外语教学, 2022(6): 78-87.
- [6] Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- [7] 韩宝成. 动态评价理论、模式及其在外语教育中的应用[J]. 外语教学与研究, 2009, 41(6): 452-458.
- [8] 教育部, 国家语言文字工作委员会. 中国英语能力等级量表[S]. 北京: 高等教育出版社, 2018.
- [9] Kosmyna, N., Hauptmann, E., Yuan, Y.T., *et al.* (2025) Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task.
- [10] Bolen, J. (2025) ESL Writing Activities, Games & Teaching Tips: Practical Ideas for the Classroom. 11-16.

附录

附录一 作文自评标准[10]

- 1) Check spelling (use your dictionary if necessary)
- 2) Check punctuation (., ; : ! ? ‘) and capital letters (BIG/small)
- 3) Grammar: Does each sentence have a subject and verb? Do they agree with each other? Check verb tense (narrative essay: past)
- 4) Check if there are run-on sentences (too long/confusing).
Run on sentence = two or more independent clauses joined without a conjunction (e.g.: He worked hard then he succeeded.)
- 5) Check if there are 2 short sentences you can combine using a conjunction: and/so/but/because, punctuation (; :), or relative clause (who, which, that)
- 6) Check if there are 1~2 words that you use too often. Can you change it?
- 7) Can you add some adjectives or adverbs to make it more interesting?
- 8) Can you add any time words or connectors to make things more clear?
- 9) Can you change some of lower-level words into something more specific/interesting?
- 10) Check the format. Did you indent the paragraph? Did you include a word-count?

附录二 同伴写作互评表

评价维度	考核标准	评价(可打√/写评语)
内容	1) 主题是否明确, 是否围绕核心观点展开? 2) 是否有吸引人的开头? 3) 内容是否具体、有细节(如案例、论据)? 4) 内容是否贴合写作目标(如读者、场景)? 5) 是否有比较好的结尾, 如用不同句式重申观点、呼吁、反问或期待等方式结尾?	
语言	1) 语句是否通顺、无明显语法/用词错误? 2) 语言风格是否匹配主题(如正式/生动)? 3) 表达是否简洁, 无冗余信息? 4) 是否有 5 个以上高级词汇? 5) 是否有 3~5 个高级句式, 如强调结构、从句、倒装、感叹句或非谓语?	
逻辑结构	1) 段落是否有清晰的中心句, 衔接是否自然? 2) 整体结构是否完整(如开头 - 主体 - 结尾)? 3) 观点/论据的排列是否有逻辑顺序(如因果、递进)?	
总的建议	请写下对这篇写作的改进建议(如补充论据、调整段落顺序等):	