

生成式人工智能辅助大学英语阅读测试文本开发研究

过君琰

(无锡学院, 江苏 无锡 214105)

摘要: 生成式人工智能的快速发展为外语测试资源的开发提供了新的技术路径。阅读是大学英语学习与语言能力评价的重要组成部分,也是大学英语四、六级考试的核心测试内容之一。因此,探索生成式人工智能辅助阅读测试资源开发具有重要的教学和测评价值。本研究考察生成式人工智能(ChatGPT)在大学英语四、六级阅读测试文本开发中的应用潜力,重点展开两类生成任务:一是以四、六级真题为参照,针对真题阅读的三种题型,分别生成相应层级的阅读测试文本;二是以大学英语教材课文为基础,将其改编成适用于四、六级三种题型的阅读测试文本。本研究从文本难度和词汇水平两个维度,将两类任务生成的文本与真题进行比较。结果表明,生成式人工智能能够基本按照目标层级和题型要求生成和改编文本,其文本难度和词汇水平与真题总体较为接近。本研究为生成式人工智能赋能外语阅读测试资源开发提供了实践启示。

关键词: 生成式人工智能; 大学英语; 阅读测试; 文本开发

基金项目: 无锡学院 2024 年“公共基础课程教学质量提升”专项课题“生成式人工智能辅助大学英语阅读测试资源库建设研究”(2024JGKT011)

DOI: doi.org/10.70693/202609285674

一、引言

随着人工智能技术尤其是生成式人工智能的快速发展,语言测试领域正在经历由传统人工开发向智能化、数字化方向转型。学者们正积极探索其在语言测试领域的应用及其机遇与挑战^{[1][2]},相关研究已触及听、说、读、写四项语言技能,但研究深度和系统性仍较为有限,主要集中在自动评分、测试反馈、测试效度、资源开发等方面^{[3][4][5][6][7]}。在这一背景下,阅读测试作为大学英语教学与测试的重要组成部分,其资源开发也面临新的应用前景和现实挑战。在大学英语四、六级考试中,阅读理解占总分的 35%,与听力理解占比相同,是分值占比最高的两类测试内容,凸显了阅读能力及其测试资源建设的重要性。生成式人工智能具有快速生成、批量处理和灵活调整等特点,可以根据测试目标、主题、体裁、篇章长度和文本难度要求辅助生成阅读文本及测试题目,因此在资源开发中展现出高效率、一致性、适应性和个性化等优势,为测试资源开发的标准化和精细化提供新的可能。如何借助生成式人工智能提升阅读测试资源的开发效率和质量,建设高质量、类型丰富、难度适切、动态更新的阅读测试资源是亟待探索的问题。本研究选取 2026 年 6 月大学英语四级和六级考试真题各 1 套以及大学英语教材课文 1 篇,运用 ChatGPT 开展阅读测试文本开发。研究首先基于四、六级真题阅读,涵盖词汇理解、长篇阅读和仔细阅读三种题型,分别生成相应的四级和六级阅读测试文本。随后基于 1 篇教材课文,按照四、六级阅读测试要求,分别改编生成包含上述三种题型的四级和六级阅读测试文本。在此基础上,通过比较生成文本与真题在文本难度和词汇水平等方面的差异,考查生成式人工智能开发不同层级、不同题型阅读测试资源的能力以及生成文本的适切性。

二、文献综述

(一) 大学英语阅读能力的重要价值和核心维度

阅读是学习者获取语言输入、积累词汇与语法知识的途径,也是听、说、写、译等技能的基础,更是培养学

作者简介: 过君琰(1984—),女,博士,研究方向为外语教学、外语测试。

通讯作者: 过君琰

生批判性思维和跨文化能力的重要载体。阅读教学是大学英语教学不可或缺的组成部分,也是大学英语测试的重要指标^[8]。综合《大学英语教学指南(2020版)》提出的阅读能力三级目标(基础目标、提高目标和发展目标)^[9]、中国英语能力等级量表(2024)中对应大学阶段的阅读理解能力(五至七级)^[10],以及《全国大学英语四、六级考试大纲(2016年修订版)》中阅读理解的考核要求与技能^[11],我们将大学英语阅读能力概括为以下几个核心维度:

1. 理解主旨信息。主要包括理解文章的主题、中心思想和篇章大意;获取重要事实、关键信息和具体细节;概括不同段落及全文的主要内容;准确检索特定信息。
2. 推断隐含意义。主要包括根据上下文推断词语、短语及句子的意义;理解文章中隐含信息和言外之意;推断作者的情感、态度和立场;理解文章的逻辑关系和深层含义。
3. 分析篇章语言。主要包括理解句子之间以及段落之间的逻辑关系,如因果、比较、条件等;运用词汇、语法及衔接手段理解篇章结构;分析文章的组织结构和论证方式;理解不同文本类型的语言特点。
4. 评析观点立场。主要包括分辨不同观点和立场;判断作者的观点、态度和立场;整合多篇相关材料的信息,对文章的语言和内容进行分析与评价;对文化现象进行批判性分析,形成自己的理解和认识。
5. 理解多类文本。阅读如英文报刊、新闻报道、社会时评等一般性材料;阅读说明书、操作手册、应用文等实用性材料;阅读专业文献、科技文章等学术型材料;阅读文学作品和文化类材料;根据不同文本类型调整理解方式。
6. 运用阅读策略。根据阅读目的、文本类型和难度选择适当的阅读方法;灵活运用略读、寻读、细读等阅读策略;通过索引等快速定位目标信息;根据需要调整阅读速度和方法。
7. 提高阅读效率。在保证理解的基础上提高阅读速度;能够顺畅地阅读篇幅较长、难度适中的英文材料。

(二) 生成式人工智能在阅读测试中的应用

生成式人工智能正在推动语言测试的设计、开发与评价发生深刻变革。尤其在阅读测试领域,生成式人工智能可以融入测试材料生成、测试试题开发以及试题资源库建设的全过程。在测试材料生成方面,生成式人工智能能够依据测试目标、语言水平、主题种类和篇章类型等生成或改编阅读文本以及调整词汇、句法和文本难度。在测试试题开发方面,生成式人工智能可以辅助生成主旨理解、细节理解、观点推断、词语猜测以及推理判断等不同层次的阅读任务,并根据要求生成选项。在试题资源库建设方面,生成式人工智能能够对不同形式的语料(如历年真题、教材文本、模拟试题等)进行整合,形成分级、分类的数字化阅读测试资源库,为优化测试和个性化学习提供有力支持。然而,目前学界对生成式人工智能在阅读测试领域的应用研究仍较为有限。

Lin & Chen (2024) 的研究以大学英语四级阅读材料和阅读子技能为基础,通过心理测量学分析和人工评价,将 ChatGPT 生成的试题与人工编写试题进行比较,探讨 ChatGPT 生成的大学英语阅读理解选择题的可行性及题目质量^[12]。研究发现,ChatGPT 基于给定材料生成的阅读选择题具有较好的可接受性,与人工编写的题目在心理测量学特征上具有一定的可比性,但是仍然存在一些局限,例如有时无法根据提示正确理解和生成选择题、偶尔答案提供不一致,以及潜在的伦理风险等。Wen & Chu (2025) 的研究主要探讨 ChatGPT 3.5 能否依据 PIRLS 2011 阅读理解框架生成高质量阅读理解题以及通过进一步提示能否改善题目质量^[13]。研究发现 ChatGPT 3.5 在生成基于文本显性信息的低阶阅读题方面表现较好,但在推断、整合和评价等高阶阅读任务中存在明显的内容错位、逻辑错误和认知过程分类问题。详细、分步的提示能够改善部分题目质量,但仍需要人工审核,且同样存在一定的伦理风险。因此生成式人工智能更适合作为阅读测试资源开发的辅助工具,能够帮助教师和学生提高试题编制的效率。罗娟、黄骞(2025)基于2023-2024年全国高考英语I卷等四套真题,运用 ChatGPT 4.0 生成相应试题,通过学生测试对比和专家结构化评估,比较生成式人工智能生成的阅读材料和人工命题材料^[14]。研究结果与 Wen & Chu (2025) 的研究类似,生成式人工智能生成的阅读材料在语言形式上已接近人工命题水平,但在题目难度、区分度与认知深度方面,尤其是高阶思维任务方面仍存在一定不足。陈大建、胡杰辉(2025)通过定性和定量相结合的方法比较了自主构建的语言测试专用大语言模型(Language assessment generative pretrained transformer, LA-GPT)和通用大语言模型(ChatGPT 4)的自动命题效果^[15]。他们的研究发现和通用大模型相比,专用模型在微技能分布和可用性评分方面展现出更多优势,在信度和拟合度方面也更接近大学英语六级阅读真题的实际测量表现。Nábělková等人(2026)探讨了人工智能生成学生阅读素养标准参照评价标准,并通过专家评价检验其质量^[16]。81.9%的 ChatGPT 5 生成评价指标获得了专家平均3分或以上(满分4分)的评价。研究结果表明,在精

确的提示工程指导下,生成式人工智能可以作为一种有用的辅助工具,用于生成与专家提出的评价标准较为相似的初步评价标准,但仍需要专家审核、教学调试和本地化校准。综合已有研究可以发现,生成式人工智能在阅读测试资源开发领域已有一定的应用,但相关研究数量仍十分有限,且存在以下的不足。首先,现有研究多集中于阅读测试题目的自动生成与编制,对阅读材料文本的生成和改编关注相对较少。其次,较少研究运用生成式人工智能对现有教材阅读文本进行改编和再开发,尚未充分实现教材内容与测试资源的有机衔接,从而更好地促进学测结合。鉴于现有研究的不足,本研究将从这两个方面展开探索。

三、研究方法

本研究选取大学英语四、六级真题阅读文本和大学英语教材阅读文本作为基础生成素材。四、六级真题阅读文本为2026年6月阅读真题(第1套),题型为词汇理解(1篇)、长篇阅读(1篇)和仔细阅读(只选择第一篇)。四级真题阅读文本内容分别为工业革命对工人阶级生活水平的影响(词汇理解)、新西兰大巴里尔岛的星空(长篇阅读)和有机食品利弊权衡之选(仔细阅读)。六级真题阅读文本内容分别为住房与社会问题租金管制(词汇理解)、数字游民(长篇阅读)、社交媒体与青少年心理健康(仔细阅读)。四级词汇理解的篇章长度为200-250词,长篇阅读的篇章长度约为1,000词,仔细阅读的篇章长度为300-350词^[11]。六级词汇理解的篇章长度为250-300词,长篇阅读的篇章长度约为1,200词,仔细阅读的篇章长度为400-450词^[11]。教材阅读文本选自《全新版大学进阶英语综合教程(第二版)》第三册第一单元Reading & Interacting部分的文章“How My Working Holiday Changed Me”。文章的主题为打工度假^[17]。

本研究的研究步骤如下:

首先运用GPT-5.6 Sol通过精细化提示词设计和优化,进行阅读文本的可控生成。研究分别基于上述四、六级真题文本和教材文本生成符合不同级别、不同题型要求的阅读语篇,即基于四级真题生成符合四级阅读要求的三种类型的阅读文本,基于六级真题生成符合六级阅读要求的三种类型的阅读文本;同时,以教材文本为基础,分别生成符合四级和六级阅读要求的三种类型的阅读文本。

其次,采用英文分级指难针先对生成语篇的整体文本难度进行测评,再考查生成文本的词汇是否符合四、六级要求^[18]。运用Vocabprofilers (<https://www.lexutor.ca/vp/comp/>)中的British National Corpus (BNC)和Corpus of Contemporary American English (COCA)对生成文本的词频进行分析,以评估生成语篇的词汇难度与适切性。在阅读理解中如果读者能够读懂文本中的98%的词汇,即98%的词汇覆盖率则被认为是理想的,读者不需要借助工具理解该文本^[19]。

最后重复上述生成与评估步骤,并根据评估结果持续优化提示词和生成文本,直至文本难度和词汇水平符合相应级别的要求。

四、研究结果和讨论

(一) 基于真题生成的阅读文本

1. 四级阅读文本: 词汇理解

表1呈现了四级真题和GPT-5.6 Sol生成的词汇理解题型文本难度和词汇水平。英文分级指难针中的文本难度等级与《中国英语能力等级量表》(2024)中的三至七级衔接。其中,3.00—3.99对应中考阶段,4.00—4.99对应高考阶段,5.00—5.99对应大学英语四级阶段,6.00—6.99对应大学英语六级阶段,7.00—7.99对应考研阶段。由表1可知,生成式人工智能产出的词汇理解文本与真题一样处于大学英语四级水平,文章单词总数符合四、六级考试大纲的要求,在200-250词之间。四级词汇能够覆盖产出文本,且最高频的3,000词(前3,000词)能够覆盖99.2%的文本词汇。

表1 四级词汇理解题型文本难度和词汇水平

阅读文本	文本难度	单词总数	四级词汇覆盖率	词频覆盖率
四级真题	5.79	243	99.59%	前3,000词覆盖98.8%
ChatGPT	5.98	236	100%	前3,000词覆盖99.2%

2. 四级阅读文本：长篇阅读

表2结果显示,生成式人工智能产出的长篇阅读同样也是大学英语四级水平,文章单词总数符合四、六级考试大纲的要求,在1,000词左右。四级词汇也能够基本覆盖产出文本,且前5,000词汇能够覆盖99.6%的文本词汇。

表2 四级长篇阅读题型文本难度和词汇水平

阅读文本	文本难度	单词总数	四级词汇覆盖率	词频覆盖率
四级真题	5.74	1,023	99.13%	前5,000词覆盖98.8%
ChatGPT	5.54	1,113	99.38%	前5,000词覆盖99.6%

3. 四级阅读文本：仔细阅读

和之前的两个题型相似,生成式人工智能产出的阅读文本在单词总数和文本难度方面与真题比较接近。由于本研究所选取的该套真题的这篇仔细阅读词汇较难,四级词汇只覆盖了该文本96.58%的词汇使用,相应地生成式人工智能产出的词汇也可以设置比前两个阅读题型的词汇更困难,但四级词汇依然能覆盖98%的文本词汇(表3)。

表3 四级仔细阅读题型文本难度和词汇水平

阅读文本	文本难度	单词总数	四级词汇覆盖率	词频覆盖率
四级真题	5.92	348	96.58%	前5,000词覆盖98.7%
ChatGPT	5.89	347	98.86%	前5,000词覆盖99.2%

4. 六级阅读文本

表4展示了六级阅读三种题型的文本难度和词汇水平。结果显示,生成式人工智能可以产出和真题文本相似难度和词汇水平的文本。六级考试词汇覆盖率均超过98%,且前5,000词汇覆盖率也超过98%。

表4 六级阅读题型文本难度和词汇水平

文本类型	阅读文本	文本难度	单词总数	六级词汇覆盖率	词频覆盖率
词汇理解	六级真题	6.16	291	99.31%	前5,000词覆盖98.6%
	ChatGPT	6.12	289	99.31%	前5,000词覆盖99.3%
长篇阅读	六级真题	6.01	1,203	97.56%	前5,000词覆盖97.3%
	ChatGPT	6.25	1,243	99.44%	前5,000词覆盖99.0%
仔细阅读	六级真题	6.81	447	99.78%	前5,000词覆盖99.3%
	ChatGPT	6.92	446	98.66%	前5,000词覆盖98.7%

(二) 基于教材改编生成的阅读文本

从表5可知,生成式人工智能(ChatGPT)能够基于教材文本改编生成符合大学英语四、六级考试要求的阅读文本。其文本难度和词数均处于相应考试的要求范围内。在词汇覆盖率方面,四、六级考试词汇对生成文本的覆盖率均达到98%以上,其中前3,000高频词汇对四级阅读文本的覆盖率达到98%以上,前5,000高频词汇对六级阅读文本的覆盖率达到98%以上,表明生成文本在词汇难度方面与相应考试要求基本匹配。

表5 教材改编阅读文本难度和词汇水平

文本类型	阅读文本	文本难度	单词总数	四、六级词汇覆盖率	词频覆盖率
词汇理解	四级	5.54	245	四级覆盖100%	前3,000词覆盖99.2%
	六级	6.02	285	六级覆盖98.6%	前5,000词覆盖98.6%
长篇阅读	四级	5.64	1084	四级覆盖99.72%	前3,000词覆盖99.4%
	六级	6.12	1168	六级覆盖99.06%	前5,000词覆盖99.4%
仔细阅读	四级	5.59	335	四级覆盖99.1%	前3,000词覆盖98.5%
	六级	6.79	447	六级覆盖98.66%	前5,000词覆盖99.6%

本研究的结果显示,ChatGPT基本能够按照大学英语四、六级阅读测试的要求生成相应的阅读测试文本,在

篇章长度、文本难度和词汇水平等方面总体表现出较好的适切性。在本研究考察的两种应用情境中,即以四、六级真题为参照生成同层级的模拟阅读文本以及将同一篇教材课文开发不同难度层级的阅读测试文本,ChatGPT均表现出较强的文本生成、难度调整和题型转换能力。这表明,生成式人工智能能够在一定程度上突破传统阅读资源开发周期较长、更新速度较慢等局限,为大学英语阅读测试资源的持续开发提供技术支持。

首先,生成式人工智能可以拓展阅读测试材料的来源。生成式人工智能可以对社会热点、科技发展、文化交流等兼具时代性和教育价值的英文文章进行改写和调整,从而补充和更新阅读测试资源,增强材料的时效性和多样性。其次,生成式人工智能还可以促进现有教学资源向测试资源转化,为实现学测结合提供了新的路径和方式。在教授一个单元的教材课文内容之前或之后,教师可以以该单元的课文为基础,根据四、六级考试的题型和难度要求,将该单元的课文改编为四、六级考试的词汇理解、长篇阅读和仔细阅读等不同类型的测试材料让学生去完成。如果是授课之前做,可以让学生提前预习授课课文,如果是授课之后再,则可以检测学生对课文的掌握情况 and 理解程度。两种方式都可以使课堂教学内容与标准化考试要求形成有效衔接。与直接使用历年真题相比,这样将日常教学材料转化为不同层级和不同题型的阅读测试资源,既可以保持教学内容的延续性,又可以帮助学生在熟悉的主题和语境中有针对性地进行训练,实现语言学习、内容知识获取、阅读能力培养与考试准备的有机结合,进一步推动教、学、评一体化。

五、结论

本研究主要探讨了 ChatGPT 在大学英语四、六级阅读测试资源开发中的文本生成与改编能力,具体包括两个方面。第一,以大学英语四、六级考试各 1 套真题的阅读部分为参照,分别针对词汇理解、长篇阅读和仔细阅读三种题型(各 1 篇),通过精细化提示词设计,引导 ChatGPT 生成与四级和六级水平相匹配的阅读测试文本。第二,选取一篇大学英语教材课文作为基础材料,在保留核心主题和主要内容的基础上,运用 ChatGPT 将文本改编生成分别适用于四级和六级三种阅读题型的文本。在此基础上,本研究从文本难度和词汇水平两个方面对上述两种方式生成的阅读测试文本进行分析,并分别与四、六级真题文本进行比较。

总体而言,本研究初步验证了生成式人工智能开发大学英语四、六级阅读测试资源的可行性。研究发现,无论是以四、六级真题为参照生成同层级的模拟阅读材料,还是基于同一篇教材课文开发不同难度层级的阅读测试材料,生成式人工智能经过多轮检测和迭代调整后,生成和改编文本与目标层级之间的匹配度进一步提高,能够较好地遵循题型、篇章长度、文本难度和词汇水平等要求。研究表明,生成式人工智能为教学材料与测试资源的衔接、不同能力层级阅读资源的开发提供了新的技术路径。因此生成式人工智能不仅可以提高阅读测试资源的开发效率,还能够拓展材料来源并持续、动态地更新测试资源。

尽管研究结果初步展现了生成式人工智能在大学英语阅读测试文本开发中的应用价值,但本研究仍存在以下局限。首先是研究主要从文本难度和词汇水平两个维度对生成式人工智能开发的阅读测试材料进行了初步评价,尚未对文本内容和语言进行系统考察,也未运用生成式人工智能生成题目,并进一步考察题目与文本的匹配程度、题目测试的阅读技能等。因此目前的研究结果主要反映生成材料在文本与词汇层面难度的基本适切性。未来研究可以在现有研究的基础上建立更加系统的阅读测试材料评价框架。一方面可以引入语言测试和心理测量学方法,通过文本质量、题目难度和区分度等指标,对生成阅读文本和试题的质量进行量化分析,另一方面,可以邀请大学英语教学与语言测试专家,从文本内容的真实性、题目设计、能力考查和测试规范等领域展开专家审查,同时还应该邀请学习者进行实际测试和访谈,获取学习者的作答成绩和反馈。通过结合上述数据内容进行信效度实证检验,从而更加全面地提高生成式人工智能开发阅读资源的质量,提升其实际应用价值。

参考文献:

- [1] CHUANG P L, YAN X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions[J]. *System*, 2025, 134. <https://doi.org/10.1016/j.system.2025.103846>
- [2] 刘建达. 人工智能时代的语言测评: 机遇与挑战[J]. *现代外语*, 2024, 47(6), 859-869.
- [3] ABUDALFA S I, BARROT J. S. Generative artificial intelligence for automated writing evaluation: A systematic review of trends, efficacy, and challenges[J]. *Assessing Writing*, 2026, 68, 101041. <https://doi.org/10.1016/j.asw.2026.101041>
- [4] ANBREEN T, ÖZTURAN T, SHRESTHA P, et al. Human vs. generative artificial intelligence in writing assessment: investigating feedback alignment, score validity, and teacher agency[J]. *Educational Assessment, Evaluation and*

Accountability, 2026, <https://doi.org/10.1007/s11092-026-09486-z>

- [5] GOH C C M, ARYADOUST V. Developing and assessing second language listening and speaking: Does AI make it better?[J] *Annual Review of Applied Linguistics*, 2025, 45, 179–199. <https://doi.org/10.1017/S0267190525100111>
- [6] GUO J. Exploring GenAI-powered listening test development[J]. *Languages*, 2026, 11(1). <https://doi.org/10.3390/languages11010017>
- [7] LIU X J, WANG J, ZOU B. Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, and oral peer feedback as feedback enhancement[J]. *Journal of English for Academic Purposes*, 2025, 75, 101505. <https://doi.org/10.1016/j.jeap.2025.101505>
- [8] 张海云. 大学英语阅读诊断反馈报告应用及评价——以 G-DINA 模型为例[J]. *外语电化教学*, 2022, 39-47, 117.
- [9] 中华人民共和国教育部高等学校大学外语教学指导委员会. 大学英语教学指南 (2020 版) [M]. 北京: 高等教育出版社, 2020: 12-16.
- [10] 中华人民共和国教育部、国家语言文字工作委员会. 中国英语能力等级量表[S]. 上海: 上海外语教育出版社, 2024: 10.
- [11] 全国大学英语四、六级考试委员会. 全国大学英语四、六级考试大纲 (2016 年修订版) [EB/OL]. 2016: 2-8.
- [12] LIN Z, CHEN H. Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items[J]. *System*, 2024, 123, 103344. <https://doi.org/10.1016/j.system.2024.103344>
- [13] WEN Z, CHU S K W. Using generative AI for reading question creation based on PIRLS 2011 framework[J]. *Cogent Education*, 2025, 12(1), 2458653. <https://doi.org/10.1080/2331186X.2025.2458653>
- [14] 罗娟, 黄骞. 生成式人工智能与人工命题的差异研究: 以高考英语阅读理解命题为例[J]. *基础外语教育*, 2025, 27(4), 97-106, 112.
- [15] 陈大建, 胡杰辉. 基于大语言模型的自动化命题研究——以英语阅读理解试题为例[J]. *外语教学*, 2025, 46(2), 40–47.
- [16] NÁBĚLKOVÁ J, KOPECKÝ K, PLISCHKE J, et al. Criterion-based assessment of reading literacy with AI[J]. *Frontiers in Education*, 2026, 11. <https://doi.org/10.3389/educ.2026.1812149>
- [17] 吴晓真, 陈进. 全新版大学进阶英语综合教程 (第二版) 第三册[M]. 上海: 上海外语教育出版社, 2024: 14-16.
- [18] 金檀, 陆小飞等. (2025). 英文分级指难针[EB/OL]. 一针三库智能教研平台(languagedata.net).
- [19] NATION I S P. How large a vocabulary is needed for reading and listening?[J] *The Canadian Modern Language Review*, 2006, 63(1), 59–82.

Developing GenAI-assisted College English Reading Test Texts

Junyan Guo

Department of General Education, Wuxi University, Wuxi, China

Abstract: The rapid advancement of Generative Artificial Intelligence (GenAI) has offered a new technological approach for the development of resources for foreign language assessment. Reading is an essential component of college English learning and language proficiency assessment and constitutes a major part of the College English Test (CET). Therefore, exploring the use of GenAI in developing reading test resources has significant pedagogical and assessment value. This study investigates the potential of GenAI (ChatGPT) for developing reading test texts for CET-4 and CET-6 through two tasks. First, using authentic CET-4 and CET-6 reading texts as benchmarks, GenAI was employed to generate texts at the corresponding proficiency levels for three reading task types. Second, a text from a college English textbook was adapted into CET-4- and CET-6-level reading texts for the same three task types. The two sets of GenAI-generated texts were then compared with authentic CET texts in terms of text difficulty and vocabulary level. The results showed that GenAI was generally capable of generating and adapting reading texts according to specified proficiency levels and task requirements. The generated texts were broadly comparable to authentic CET texts in text difficulty and vocabulary level. This study offers practical implications for GenAI-powered foreign language reading test development.

Keywords: Generative Artificial Intelligence; College English; Reading test; text development