



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 7(3), 102640 (2025)

<https://doi.org/10.29140/tltl.v7n3.102640>

Spontaneous Use of Generative Artificial Intelligence in Preparing for College English Test: GenAI Adoption Factors, Learners' Motivation, and Self-Efficacy



TIANSHI HAO^a 

XINYU ZHAO^b 

ZIJIN LU^c 

^a*School of Education and Psychology,
Pepperdine University, United States
Haots2308@gmail.com*

^b*Graduate School of Human Science,
The University of Osaka, Japan
zhaoxinyu2018316@gmail.com*

^c*Department of English Language and Literature, Faculty of Arts and Social Sciences,
Hong Kong Baptist University, Hong Kong
22483217@life.hkbu.edu.hk*

Abstract

This study investigates the adoption and impact of generative artificial intelligence (GenAI) tools on English as a foreign language (EFL) learners, focusing on self-efficacy, motivation, and predicted exam performance. Using a mixed-methods approach, the research combines survey data ($N = 595$) and interviews ($N = 8$) from Chinese university students preparing for the College English Test 4. The quantitative analysis shows that positive attitudes, perceived usefulness, and ease of use predict GenAI usage frequency, while intrinsic motivation negatively correlates with use. GenAI use frequency did not significantly affect predicted test scores or self-efficacy levels. Younger students used GenAI more frequently. The qualitative data reveal that students primarily use GenAI as supplementary learning resources, which, while beneficial, do not substantially enhance self-efficacy or predicted exam performance. The study found a knowledge gap in the effectiveness of these tools in EFL contexts. It suggests strategies for optimizing GenAI in EFL learning, including blending with traditional methods,

Copyright: © 2025 Tianshi Hao, Xinyu Zhao & Zijin Lu. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

providing user training, balancing technology with traditional approaches, ensuring equal access, and customizing GenAI for education. The findings imply a need for educators to supervise GenAI use and for institutions to facilitate effective integration. Future research should examine the long-term and cultural effects of GenAI use in education.

Keywords: generative AI (GenAI), English as a foreign language (EFL), self-efficacy, motivation, technology enhanced language learning, mixed method.

Introduction

Generative artificial intelligence (GenAI) is a disruptive technology which has impacted education (Miller, 2024) and learning, including in the area of English as a foreign language (EFL) learners (Richards & Schmidt, 2013). It can benefit EFL by providing writing feedback (Li, Huang, et al., 2024), enhancing foreign language enjoyment (Xu et al., 2024), and thus increasing self-efficacy (Yildiz, 2024). In second-language research, self-efficacy refers to learners' belief that they can meet a task's demands predicts persistence and eventual proficiency beyond actual ability (Bandura, 1997). Motivation is the effort, the desire to achieve a language goal, and a positive attitude toward learning (Gardner, 1985). Motivation can reflect intrinsic interest as well as extrinsic reasons such as rewards or requirements (Noels et al., 2000). Self-efficacy and motivation can greatly affect learning outcomes and GenAI can affect both constructs, with its personalized feedback and nonjudgmental language learning environment, which interests the learner (Anjum et al., 2025) and allows low-stake rehearsal that reduces learning anxiety (Huang, 2024). GenAI tools could also motivate and increase students' interest in speaking skills (Muniandy & Selvanathan, 2024). Because the intention to keep using a tool is the behavioral outcome of motivation, the present study also adopted the Technology Acceptance Model (TAM), in which perceived usefulness and ease of use foster a favorable attitude toward a tool, thereby increasing the intention to keep using it in the future (Davis, 1989).

As confirmed by two systematic reviews, the majority of studies on the intersection of GenAI and EFL tend to investigate its intention to use and rarely use experimental methods to test GenAI in controlled settings (Balci, 2024). Within those empirical studies, most of them test the benefit and challenges of using GenAI in the EFL process and the use of GenAI in writing but “few studies have quantitatively examined its effects on students' performance and motivation” and the impact of GenAI on other aspect of language skills such as reading, speaking, and listening (Lo et al., 2024, p.1). Some studies have examined speaking self-efficacy in a controlled environment (Celik et al., 2025) and have found that GenAI tools moderate the relationship between self-efficacy and learning outcome (Mahande et al., 2025); however, the generalizability of those articles might be limited, as they recruited fewer participants with a single outcome, focused on a single tool (Celik et al., 2025), and over-relied on quantitative outcomes (Mahande et al., 2025). Using GenAI as an intervention also excludes voluntary usage, which could further limit the generalizability and explainability of the findings.

To address these gaps, the current study conducted a mixed-method study with 595 undergraduate students at various grades in China from a single university and examine how their spontaneous GenAI usage during their College English Test Band 4 (CET-4) preparation affected their learning motivation and self-efficacy. This study focuses on Chinese learners because China has more than 400 million EFL learners (Pan, 2019). The study selected the CET-4 because it is mandatory for all Chinese students to take and eventually pass (score 425 out of 710), starting in their late freshman year in our data collection site. This test was designed to align with Chinese students' learning backgrounds (Yan & Huizhong, 2006), making it a fairer assessment tool than international tests, which require additional

understanding of international culture. The CET-4 is administrated twice a year, assessing listening, reading, and writing skills. Only those who pass the CET-4 can register for CET-6 (Xie et al., 2025).

The findings could contribute to existing research by contributing additional data that combine affective factors with actual usage data. Instructors would get evidence about which learners gain confidence and which remain hesitant with GenAI, and therefore adjust their strategy when helping students prepare for the tests. University leadership would see how voluntary GenAI adoption influences self-efficacy and motivation, which may inform them as they continue to revise their campus GenAI policy. This study was guided by two research questions:

- RQ1: What factors (motivation, self-efficacy, TAM) are associated with the frequency of GenAI usage for English language learning?
- RQ2: How does the use of GenAI influence both predicted English learning outcomes and learners' self-efficacy during the test?

Literature Review

GenAI-Assisted Language Learning

GenAI facilitates independent study, such as language learning and test preparation. In Duolingo's GPT-4 "Roleplay" and "Explain My Answer" modes, 91–97% of 385 learners said the live explanations and conversation drills helped them speak, parse grammar, and learn from their mistakes (Kittredge et al., 2025). Self-learning adults also benefit from GenAI's ability to create level-matched material the moment a question appears (Li, Wang, & Bonk, 2024). However, not all students use GenAI equally frequently. Low-performing students' trust has declined over time, suggesting that AI tools may require prolonged scaffolding for struggling learners, whereas high performers experienced consistent utility (Song, 2024). An experimental study confirmed this point, where GenAI users achieved the largest essay score gains but the study also cautioned against "metacognitive laziness", which could undermine long-term learning outcomes (Fan et al., 2024). Further, a systematic review of 38 studies found that 71.4% link GenAI's instant, adaptive feedback to stronger self-regulated learning, while also cautioning that heavy dependence can blunt initiative (Sardi et al., 2025). Given the usefulness of GenAI in benefitting learning and providing additional help to struggling students, we hypothesize the following:

Hypothesis 1: Students with lower prior scores may rely on GenAI to improve their performance.

GenAI has proven to be versatile and has a great impact on education (Miller, 2024). Studies show that when GenAI is integrated into learning, it can enhance language learning outcomes and overall learning satisfaction (Jadhav et al., 2024), which could explain why there are more articles discussing the adoption of GenAI into learning. The adoption of GenAI is not merely based on the urge to improve performance but is also related to attitudes towards GenAI. For instance, Zheng et al. (2024) found that expected outcomes, effort, hedonic motivation, habit, and social influence predict Chinese EFL learners' behavioral intentions towards GenAI tools. Additionally, Li, Huang et al. (2024) extended the TAM to include cognitive and human–computer interaction factors, showing that intrinsic motivation, readiness, and interface quality impact the perceived usefulness and ease of use, thus affecting attitudes toward AI in education. Lastly, Liu et al. (2024) examined informal digital learning of English with GenAI, identifying themes of perceived usefulness and practical applications, supporting the idea that positive attitudes enhance GenAI's adoption in education. However, given the improvements and

changes in GenAI, better understanding and practices have been established in the past few years, so a new study is necessary to test if this connection stays true.

Hypothesis 2: Students with a positive attitude towards GenAI are more likely to use GenAI to learn English.

Motivation and Adoption of GenAI

GenAI remains an optional aid in most English-learning courses (Lee & Davis, 2024), given the fact that many universities worldwide do not have formal guidelines regarding GenAI (UNESCO, 2024). The very act of opening the chatbot reflects a learner's motivation. Students spontaneously use tools because they want results and not merely because GenAI is available. However, it has been found that GenAI tools could elicit intrinsic motivation and gives participants a greater feeling of choice and competence (Leong et al., 2024). When combining motivation with aspects of TAM, it has been found that intrinsic motivation and perceived usefulness were the greatest predictor for the adoption of GenAI (Huang & Mizumoto, 2024). Students' motivation to learn English converts to the intention to use technology such as GenAI (Zheng et al., 2024). Driven students keep revising prompts and comparing multiple answers, whereas classmates with lower motivation often drop out after a single shallow try or stop using GenAI for study at all (Ammari et al., 2025; Annamalai et al., 2025). Over time the difference between the learning outcomes of students with different levels of learning motivation will be evident. However, most GenAI in EFL studies treat motivation as an outcome of using the tool (Deng et al., 2025; Woo et al., 2024), but this has mostly been established in a controlled environment which limits the ecological validity. Therefore, use motivation as predictor for usage, tested in a naturalistic environment is worth exploring.

Hypothesis 3: Students with stronger motivation to learn English, regardless of the type of motivation, use GenAI more often.

GenAI and Self-Efficacy

The integration of GenAI into the classroom has increased students' confidence within different subjects. For instance, one study showed that EFL students who swapped conventional speaking drills for a GenAI tutor scored higher in their speaking self-efficacy (Celik et al., 2025). Agriculture undergraduates who used ChatGPT to generate Arduino code were observed to have a statistically significant increase in their programming self-efficacy after a single laboratory session (Johnson et al., 2024). A similar result was reported in other computer science courses, where students report higher self-confidence once ChatGPT had guided their coding practice (Yılmaz & Yılmaz, 2023). Translation majors who treated ChatGPT as a partner likewise rated their own performance higher and handled tasks with less anxiety (Andrade Preciado & González Vallejo, 2025). It appears that after success with the tool, students judge GenAI as more useful and intend to keep it in regular use (Johnson et al., 2024).

It is clear that GenAI can increase learners' self-efficacy. This study collected predicted scores as a supplement to self-efficacy, which served as a measurement for self-efficacy during the exam because fewer studies have directly addressed English learners' predicted score and their self-efficacy. Other studies compared self-evaluations with objective proficiency in learning other languages to observe how learning attitude can affect self-ratings (Hernández-Rivera et al., 2024). Therefore, two hypotheses are proposed, where Hypothesis 4 directly aims to address the abovementioned gap. We hypothesized that participants in this study who use technology such as GenAI more frequently would also have higher

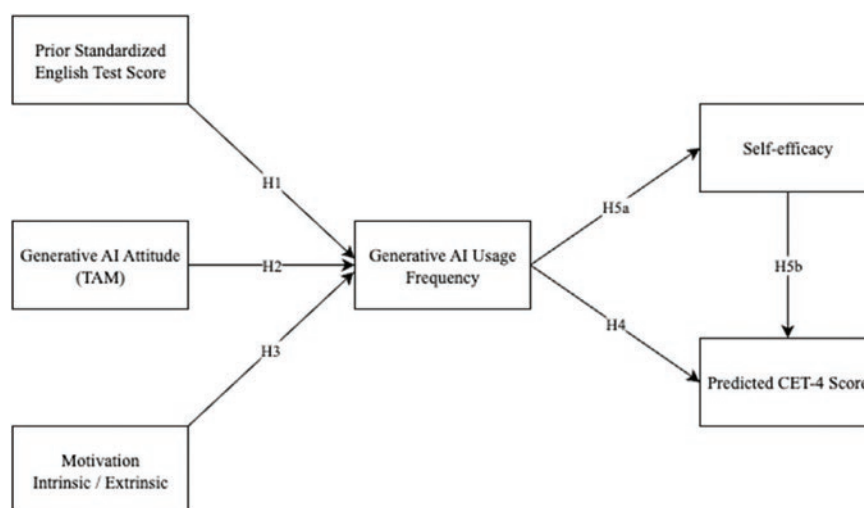


Figure 1 Variables' Relationships and Hypotheses.

self-efficacy. A summative figure is provided above to display variables' relationships and hypotheses used in this study (Figure 1).

Hypothesis 4: GenAI usage could lead to better predicted scores.

Hypothesis 5: GenAI usage boosts self-efficacy related to learning English (5a), influencing the prediction of improved English score (5b).

Method

Research Design

This study followed an explanatory sequential mixed-methods design, starting with quantitative data collection, followed by gathering qualitative data to deepen our understanding of the results (Shorten & Smith, 2017). This approach addresses the limitations of both methods, providing comprehensive insights into the studied population (Dawadi et al., 2021). In the first phase, we examined the relationships among the factors in light of our hypotheses, and the second phase provided in-depth qualitative insights to supplement the quantitative findings. The qualitative part using quantitative ethnography, a method that combines ethnographic and statistical techniques to analyze patterns of human interaction (Shaffer, 2017). The study received institutional review board approval, and all participants provided informed consent before participating.

Participants and Sampling

Power Analysis. We performed a priori power analyses using G*Power 3.1.9.7 (Mayr et al., 2007) to ensure adequate power for detecting significant effects. According to Hypotheses 4 and 5, the calculation involved 15 predictors from Table 1, excluding the attention-check question. With an effect size (f^2) of 0.15, an alpha level of 0.05, and a desired power ($1 - \beta$) of 0.80, at least 139 participants were required. Because of GenAI's limited access in China and uncertainty about its popularity among Chinese students, we aimed to gather as much data as possible to avoid a small sample size of AI users in English exam preparation.

Table 1 *Demographic Characteristics of the Survey Respondents*

Characteristic	Mean (SD) or <i>n</i> (%)
Sample size, <i>N</i>	595
Age (years)	19.84 (1.11)
Female	363 (61.0%)
Male	201 (33.8%)
Third gender/no answer	6 (1.0%)

Note. Percentages are calculated from the total sample (*N* = 595); they may not sum to 100% because of rounding. SD, standard deviation.

Quantitative Data Collection. We used a nonprobability sampling technique (MacNealy, 1999), selecting participants on the basis of availability and willingness. We initially collected 896 responses from university students in Shandong Province, China. The dataset was cleaned using three attention-check questions: two required a specific answer to be selected, and another compared gender responses at the beginning and end for consistency. Participants needed to pass two of three checks to be included, resulting in a final dataset of 595 participants (see Table 1).

Qualitative Data Collection. This study contacted 35 students who indicated they would participate in the interview and provided correct contact information. The final dataset involved eight undergraduate students enrolled in different majors at different stages of their study; participants were categorized by their experience with GenAI tools. All participants are Chinese native speakers (see Table 2). The qualitative data was collected with 10 questions via a semi-structured interview online and were completed one week after the survey. Recordings were transcribed by hand and checked line-by-line by at least two researchers. Each transcript was anonymized and labeled P1–P8 before analysis (see Appendix A for the interview questions).

Instruments

Survey. The survey in this study was structured according to the appropriate theoretical frameworks. For self-efficacy, we focus on academic exam self-efficacy, with scales adapted from the Academic Exam Self-Efficacy (Nielsen et al., 2017) and Bandura's guidelines (2006). To assess motivation,

Table 2 *Academic Characteristics of the Interviewees*

Participant	Year	Major	Usage status
P1	Sophomore	Computer Science and Technology	User
P2	Senior	Mathematics and Applied Mathematics	User
P3	Senior	Finance	Nonuser
P4	Junior	Statistics	User
P5	Sophomore	Cultural Industry Management	User
P6	Senior	Finance	Nonuser
P7	Junior	Computer Science and Technology	User
P8	Senior	Finance	Nonuser

we adapted the English Language Learner Motivation (Ardasheva et al., 2012) and the Academic Motivation Scale (Vallerand et al., 1989) to explore undergraduate students' motivation for English learning. For GenAI tool acceptance, we applied factors from the TAM3, the most updated and comprehensive version of TAM, covering perceived usefulness, perceived ease of use, attitudes, and behavioral intention. In total, the survey comprised 63 questions (Table 3).

Table 3 *Structure and Factors of the Surveys*

Theory/Scales	Factors	Type	Sample items (number)
Self-Efficacy (Bandura, 2006; Nielsen et al., 2017)	Listening Test (LST)	Likert scale (1–5)	I can grasp the main points and details of the recording. (3 items)
	Reading test (RST)	Likert scale (1–5)	I can understand the themes and details of reading paragraphs. (3 items)
	Writing test (WRT)	Likert scale (1–5)	I can write coherent and logical articles in English. (3 items)
	Translation test (TRS)	Likert scale (1–5)	I can accurately translate sentences from Chinese to English, using correct grammar and vocabulary. (3 items)
	Overall exam skills (EXM)	Likert scale (1–5)	I can effectively manage time across different parts of the exam. (3 items)
Motivation (Ardasheva et al., 2012; Deci & Ryan, 1985; Gardner & Lambert, 1972; Vallerand et al., 1989)	Intrinsic motivation (ENG_IN)	Likert scale (1–5)	I enjoy the challenge and satisfaction of learning new vocabulary and grammar. (10 items)
	Extrinsic motivation (ENG_EX)	Likert scale (1–5)	Proficiency in English helps me achieve my career goals. (10 items)
Technology acceptance (Davis, 1989; Venkatesh & Bala, 2008)	Perceived usefulness (AIU)	Likert scale (1–5)	I find that using AI tools has improved my learning efficiency. (5 items)
	Perceived ease of use (AIE)	Likert scale (1–5)	I can use AI tools to perform unfamiliar tasks without guidance. (5 items)
	Behavioral intention (BHI)	Likert scale (1–5)	I plan to use AI tools for other exams in the future. (5 items)
	Attitude (ATT)	Likert scale (1–5)	Using AI tools is helpful for exam preparation. (5 items)
	Frequency of use (AIF)	Likert scale (1–6)	How often do you use artificial intelligence (GenAI, AI) to help prepare for exams? (1 item)
Other questions	Previous score (Prvs_Eng)	Likert scale (1–5)	What is your previous Gaokao English score? (1 item)
	Predicted score (SCR)	Likert scale (1–5)	How would you predict your next CET-4 score? (1 item)
	Gender (asked twice)	Multiple choice	What is your gender? (2 items)
	Age	Numerical	What is your age? (1 item)
	Attention check	Multiple choice	To demonstrate attention, please select <i>strongly disagree</i> . (2 items)

Data Analysis

Quantitative Data Analysis. The data analysis was conducted using SPSS. We used multiple linear regression to test Hypotheses 1 to 5. In addition, we also expected to use mediation analysis to further verify Hypothesis 5 because Hypothesis 5 predicted that self-efficacy is the reason why GenAI usage frequency affects predicted English scores.

Qualitative Data Analysis. Interview data were transcribed and anonymized, and each participant was assigned a unique identifier. To ensure accuracy and consistency, two researchers independently coded the responses, maintaining inter-rater reliability and resolving discrepancies (Armstrong et al., 1997). An inductive coding process was used to allow themes to emerge (Thomas, 2006), resulting in one codebook (see Appendix B) aligned with each research question. Epistemic network analysis (ENA) was used to create a visual and statistical representation of how EFL learners connect perceptions of GenAI with their past scores, motivation, and self-efficacy. ENA models patterns of code connections in data, assuming that researchers can consistently define a meaningful set of features, that the data are organized into local conversational units, and that the analytic focus is on how those features link within each unit (Shaffer, 2017). The ENA records whether each pair of codes co-occurs, then sums those binary line-level networks to construct an aggregate model for each analytic unit (Ruis et al., 2019; Siebert-Evenstone et al., 2017). Because the responses were separated by participant and by question, we chose a window size of 1 to reflect our data structure. A minimum edge weight of 0.05 and a scaling factor of 1.5 were applied to enhance visibility and emphasize the data's most significant connections. Models were generated to compare insights between those who use GenAI more frequently and those who use it less or do not use it at all according to their responses.

Results

Quantitative Results

Descriptive Statistics. The 595 participants in our study were all college students from China, with an average age of 19.84 years (standard deviation (SD) = 1.11). Of these, 363 were female, 201 were male, and 6 self-identified as third gender or did not answer the gender question (see Table 4).

Participants scored between 110 and 129 on the college entrance English exam, with a positive sense of self-efficacy for the CET-4 test, though their listening self-efficacy was lower. Participants generally viewed GenAI positively for learning English, finding it useful, easy to use, and likely to assist in future exams. They reported strong motivation to learn English, with external motivations (e.g., career prospects) outweighing internal ones. Most students predicted passing the CET-4 exam.

Cronbach's alpha and factor loadings (see Table 4 and Table 5) confirmed good reliability for self-efficacy, GenAI acceptance, and motivation factors. Therefore, we averaged the scores within each factor for subsequent analysis. Because of low factor loadings and communalities in the ENG scale, we conducted further factor analysis, leading to the exclusion of certain items. Table 5 lists the items retained for analysis, along with their factor loadings and communalities.

Hypothesis Testing. We combined Hypotheses 1 to 3 into a regression model to predict factors related to the frequency of GenAI use in English learning (RQ1; see Table 6). A multiple regression analysis was conducted with age, gender, past English scores (Prvs_Eng), GenAI acceptance (perceived usefulness (AIU), perceived ease of use (AIE), behavioral intention (BHI), attitude (ATT)), and English learning motivation (internal (ENG_IN) and external (ENG_EX)) as independent variables, and GenAI use frequency (AIF) as the dependent variable. The model showed a medium effect size ($R^2 = 0.167$).

Table 4 *Descriptive Statistics and Reliability of Each Factor*

	Mean	Median	SD	α
Prvs_Eng	2.88	3.00	0.70	–
LST	2.91	3.00	1.11	0.84
RST	3.68	4.00	0.87	0.91
WRT	3.27	3.33	0.90	0.90
TRS	3.12	3.00	0.90	0.90
EXM	3.55	3.67	0.87	0.88
AIF	3.15	3.00	1.55	–
AIU	3.41	3.40	0.78	0.88
AIE	3.34	3.40	0.59	0.63
BHI	3.43	3.40	0.71	0.84
ATT	3.64	3.80	0.73	0.93
ENG_IN	3.24	3.29	0.97	0.94
ENG_EX	3.88	4.00	0.80	0.92
SCR	3.94	4.00	0.93	–

Note: Abbreviations are explained in Table 3.

Table 5 *Factor Loadings of Each Item in the Motivation (ENG) Scale*

Items	Factor loadings		Communality
	Intrinsic motivation	Extrinsic motivation	
ENG_8	0.90		0.75
ENG_3	0.87		0.74
ENG_7	0.85		0.70
ENG_2	0.82		0.64
ENG_4	0.81		0.66
ENG_5	0.78		0.67
ENG_15	0.67		0.60
ENG_18		0.85	0.70
ENG_14		0.82	0.64
ENG_17		0.80	0.66
ENG_13		0.73	0.51
ENG_10		0.73	0.61
ENG_9		0.71	0.57
ENG_19		0.89	0.77

Note: Factor loadings < 0.32 are suppressed. Communality scores < 0.45 are removed.

Table 6 Regression Analysis for Research Question 1

	B	SE	t	95% CI	p-value
Intercept	5.863	1.190	4.926	[3.525, 8.201]	0.000
Age	-0.277	0.054	-5.101	[-0.384, -0.171]	0.000
Gender	0.084	0.121	0.693	[-0.154, 0.322]	0.489
Prvs_Eng	-0.011	0.091	-0.119	[-0.189, 0.167]	0.905
AIU	0.351	0.106	3.317	[0.143, 0.559]	0.001
AIE	0.291	0.115	2.520	[0.064, 0.517]	0.012
BHI	0.289	0.133	2.177	[0.028, 0.550]	0.030
ATT	0.012	0.130	0.093	[-0.244, 0.268]	0.926
ENG	-0.176	0.075	-2.353	[-0.323, -0.029]	0.019
ENG	0.046	0.089	0.517	[-0.130, 0.222]	0.606

The analysis revealed that younger students were more likely to use GenAI for learning English ($B = -0.27$, $t = -5.10$, 95% confidence interval (CI) $[-0.384, -0.171]$, $p = 0.000$). Gender and past English scores were not significant predictors. GenAI use was higher among those who found AI useful (AIU: $B = 0.35$, $t = 3.32$, 95% CI $[0.143, 0.559]$, $p = 0.001$), easy to use (AIE: $B = 0.29$, $t = 2.52$, 95% CI $[0.064, 0.517]$, $p = 0.012$), and those who planned to use AI in the future (BHI: $B = 0.29$, $t = 2.18$, 95% CI $[0.028, 0.550]$, $p = 0.030$). General attitudes toward AI did not significantly affect usage frequency. For learning motivation, higher intrinsic motivation reduced GenAI use frequency ($B = -0.18$, $t = -2.35$, 95% CI $[-0.323, -0.029]$, $p = 0.019$), while extrinsic motivation had no significant effect.

We used a multiple regression model to test Hypotheses 4 and 5, exploring the factors affecting predicted scores (RQ2; see Table 7). Independent variables included age, gender, past English scores (Prvs_Eng), GenAI usage frequency, self-efficacy in English tests (listening test (LST), reading test (RST), writing test (WRT), translation test (TRS), overall exam skills (EXM)), AI acceptance (AIU, AIE, BHI, ATT), and motivation for learning English (ENG_IN, ENG_EX). The dependent variable was the students predicted English test results. The model had a high effect size ($R^2 = 0.505$).

The analysis showed that the frequency of GenAI use in test preparation did not significantly impact predicted test scores ($B = -0.013$, $t = -0.110$, 95% CI $[-0.247, 0.220]$, $p = 0.913$). Age, gender, past scores, and AI acceptance also did not significantly influence predictions. However, students with high overall self-efficacy (EXM) predicted better test outcomes ($B = 0.602$, $t = 2.48$, 95% CI $[0.126, 1.079]$, $p = 0.013$). Both intrinsic and extrinsic motivation significantly predicted higher exam scores (ENG_IN: $B = 0.33$, $t = 8.86$, 95% CI $[0.260, 0.408]$, $p = 0.000$; ENG_EX: $B = 0.41$, $t = 12.13$, 95% CI $[0.342, 0.474]$, $p = 0.000$). Because the frequency did not affect score predictions, we did not pursue mediation analysis for Hypothesis 5.

However, we used simple regressions to explore the relationship between the frequency of GenAI use and self-efficacy (LST, RST, WRT, TRS, EXM) as predicted in Hypothesis 5a (see Table 8). We found that in this survey, increased use of GenAI in exam preparation (AIF) did not improve students' self-efficacy (LST, RST, WRT, TRS, EXM).

Table 7 *Regression Analysis for Research Question 2*

	B	SE	t	95% CI	p-value
Intercept	4.006	0.560	7.153	[5.106, 7.153]	0.000
Age	−0.009	0.026	−0.341	[−0.060, 0.042]	0.733
Gender	0.076	0.058	1.305	[−0.038, 0.189]	0.192
Prvs_Eng	−0.064	0.044	−1.438	[−0.151, 0.023]	0.151
AIF	−0.013	0.119	−0.110	[−0.247, 0.220]	0.913
LST	−0.120	0.159	−0.757	[−0.432, 0.192]	0.449
RST	0.404	0.260	1.551	[−0.107, 0.915]	0.121
WRT	−0.095	0.269	−0.355	[−0.623, 0.432]	0.723
TRS	−0.241	0.266	−0.904	[−0.764, 0.282]	0.366
EXM	0.602	0.243	2.483	[0.126, 1.079]	0.013
AIU	−0.222	0.252	−0.882	[−0.717, 0.272]	0.378
AIE	−0.488	0.271	−1.801	[−1.020, 0.044]	0.072
BHI	0.197	0.310	0.637	[−0.412, 0.807]	0.525
ATT	0.235	0.308	0.762	[−0.371, 0.841]	0.446
ENG_IN	0.334	0.038	8.859	[0.260, 0.408]	0.000
ENG_EX	0.408	0.034	12.127	[0.342, 0.474]	0.000

Table 8 *Regression Analysis for Hypothesis 5a*

Self-efficacy		B	SE	t	95% CI	p-value
LST	Intercept	0.593	0.021	28.632	[0.552, 0.634]	0.000
	AIF	−0.021	0.035	−0.605	[−0.091, 0.048]	0.546
RST	Intercept	0.756	0.016	46.920	[0.724, 0.788]	0.000
	AIF	−0.038	0.028	−1.374	[−0.092, 0.016]	0.170
WRT	Intercept	0.650	0.017	38.841	[0.618, 0.683]	0.000
	AIF	0.008	0.029	0.278	[−0.048, 0.064]	0.781
TRS	Intercept	0.618	0.017	36.861	[0.585, 0.651]	0.000
	AIF	0.013	0.029	0.461	[−0.043, 0.069]	0.645
EXM	Intercept	0.717	0.016	44.449	[0.685, 0.748]	0.000
	AIF	−0.011	0.028	−0.409	[−0.065, 0.043]	0.683

Qualitative Results

Three ENA graphs were generated for both of the research questions. Different nodes represent the spaces. The lines indicate connections between constructs, with the thickness corresponding to the strength of that connection. The ENA graph for GenAI user and nonusers and their subtracted graph are shown in Figure 2.

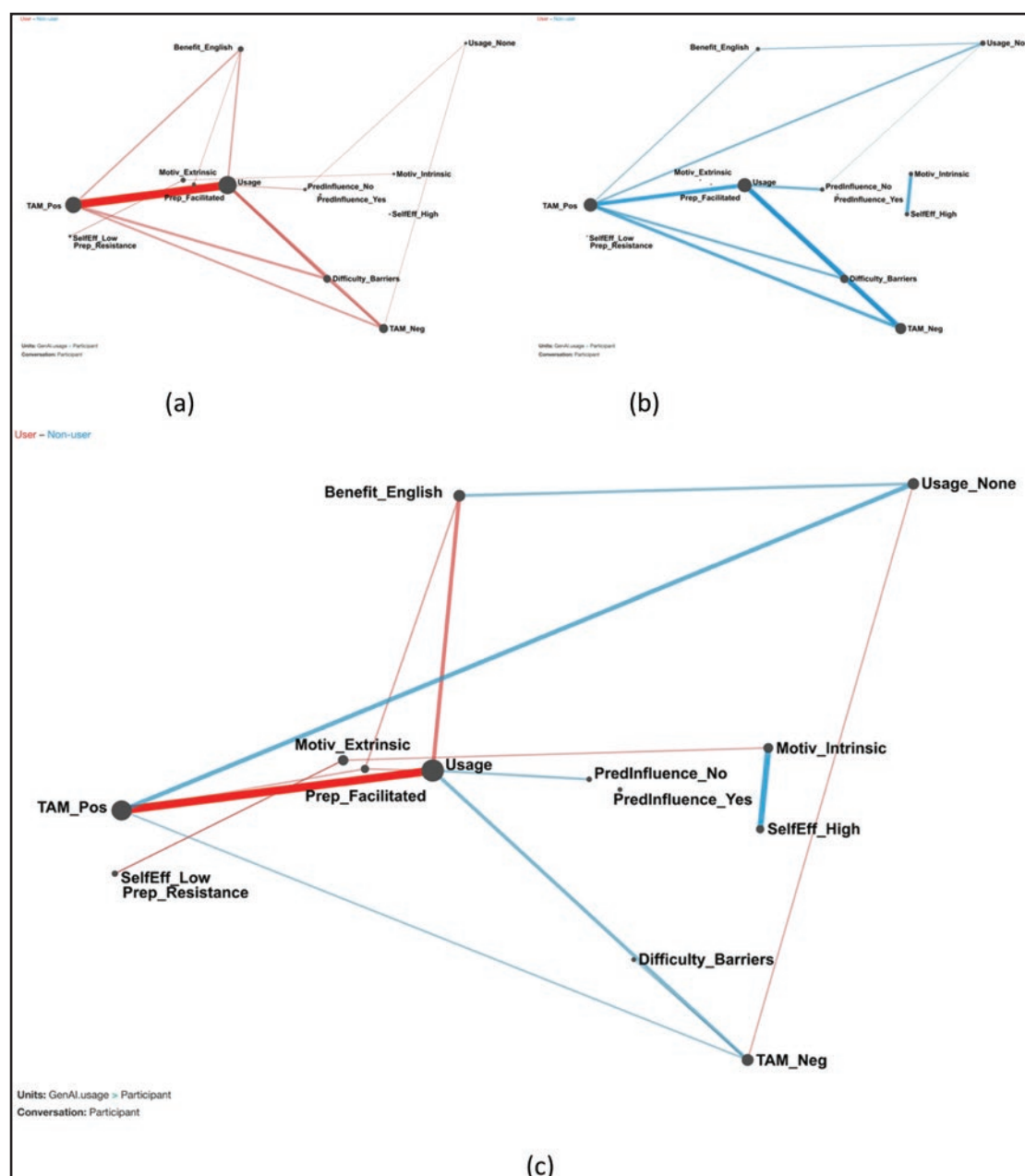


Figure 2 ENA Graphs.

Note. The red lines represent users' responses, and the blue lines represent nonusers' responses. The third subtractive graph shows a direct comparison between the two groups.

When examining the subtracted model, “TAM_Pos” has the strongest connections with “Usage” (P1: “Regarding AI, it mainly helps me clarify professional knowledge in my current studies”). The links between “Motiv_Intrinsic” and “SelfEff_High” (P4: My genuine passion for English makes me actively notice exam-related content, song releases, and musical revivals. Lyrics/dialogues contain excellent vocabulary/sentence patterns helps in CET-4 exams and deepens question comprehension), “TAM_Pos” and “Usage_None” (P3: I don’t think they’re related [motivation to learn English and my attitude toward AI]. But AI can promote English learning), “Usage” and “TAM_Neg” (P2: [I use AI] very frequently, for learning and programming... AI software sometimes fails to clearly recognize questions or instructions, requiring repeated adjustments... has many ads, and its answers

are imprecise or even incorrect. After... I have no desire to use AI to improve my English, but I still force myself to study English [because it is important for career]) also showing relatively strong connections with each other. In contrast, the weakest connection lies between “Motiv_Extrinsic” and “Prep_Resistance” (P2: My intrinsic interest is low... I’m not drawn to languages... I feel resistant during exam prep; lacking interest and aptitude, I forget words instantly. During practice, I recognize words but can’t recall their meanings), “Prep_Facilitated” and “benefit_English” (P1: While preparing for CET-4, I encountered many concepts I didn’t understand, so I used AI to assist my learning. For instance, when grappling with English grammar nuances or context-specific word meanings, AI provided examples from global sources to clarify. Mainly for writing, I had AI generate an English essay, then I scored its work. Later, I submitted my own essay for AI evaluation), and “TAM_Neg” and “Usage_None” (P2: Minimal impact. CET exams mainly tests vocabulary and listening skills, where AI is rarely or never used). Along the X axis, a two-sample t-test assuming unequal variance showed that users (mean = -0.27 , SD = 0.17 , $N = 4$) were statistically significantly different at the $\alpha = 0.05$ level from nonusers (mean = 0.27 , SD = 0.172 , $N = 4$; $t(6.00) = -4.38$, $p = 0.00$, $d = 3.10$).

Discussion

GenAI has revolutionized traditional teaching methods and brought new opportunities and challenges to EFL learners. This study aimed to explore the influence of GenAI tools in EFL learning. In a natural context, we used Chinese college students preparing for the CET-4 English test as the subjects, and designed and verified five hypotheses for the following two research questions.

Research Question 1: What Factors are Associated with the Frequency of GenAI Utilization for English Language Learning?

We predicted that students with lower past scores (Hypothesis 1), positive attitudes toward AI (Hypothesis 2), and strong motivation to learn English (Hypothesis 3) would use GenAI more frequently in their test preparation. However, contrary to Hypothesis 1, students’ prior English scores showed no significant relationship with the frequency of adopting GenAI. Hypothesis 2 was partially supported. Learners who perceived GenAI as useful or easy to use utilized it more frequently, aligning with the TAM (Davis, 1989). Qualitative data supported this, with students linking their experiences with GenAI to both positive and negative outcomes.

Unexpectedly, Hypothesis 3 revealed that students with higher intrinsic motivation to learn English were less likely to use GenAI in test preparation. This negative association suggests that highly intrinsically motivated learners prefer to stay with methods they can control and choose resources that match their learning goals (Hou, 2024; Ryan & Deci, 2000). In other words, learner agency significantly influences language learning (Jang, 2022). Studies on learning technology have found that intrinsic motivation increases adoption only indirectly by strengthening perceived usefulness rather than directly increasing behavioral intention (Sun & Gao, 2020). This pattern suggests that autonomy-oriented learners adopt technology selectively, using it as an enhancement rather than a core resource (Liu et al., 2016).

Younger students were more inclined to use GenAI for test preparation, consistent with the findings of Chan and Lee (2023), who found that younger teachers are more proficient with new technologies. An experimental study in Denmark showing that every incremental year was associated with a 0.6–0.7% lower likelihood of adopting ChatGPT for work (Humlum & Vestergaard, 2025). Our result suggests a generational gap in technology adoption, with younger individuals being more comfortable and frequent users of tools like GenAI. However, qualitative insights indicate that nonuse among older students may be more related to individual differences than age.

Research Question 2: How Does the Use of AI Influence Both English Language Learning Outcomes and Learners' Self-Efficacy in The Language Learning Process?

Our data did not support Hypothesis 4. The regression analyses showed no statistically significant link between the frequency of GenAI use and students' predicted CET-4 grades. The subtracted ENA graph is consistent with this result because usage is strongly connected only to "TAM_Pos" (usefulness and easy to use of GenAI) and "Benefit_English" (GenAI is beneficial to English learning); it displays no edge link to any motivational or self-efficacy code. This suggests that students may not perceive GenAI as directly enhancing their test readiness. This disconnect might be attributed to the lack of formal introduction to GenAI tools, as they are still emerging technologies and students are still learning to integrate them effectively (Kurtz et al., 2024; Walter, 2024). Our findings also reflect the influence of personality traits, digital barriers, and the poor quality of available GenAI tools, which limit their usefulness in English learning.

Since Hypothesis 4 was unsupported, the expected positive impact of AI use on self-efficacy (Hypothesis 5) was not observed. Nonetheless, we explored whether the use of GenAI would affect students' self-efficacy (Hypothesis 5a). Increased GenAI usage did not significantly enhance students' self-efficacy for English. This could be caused by unfamiliarity with GenAI, restricted access, and learning styles. Similarly, Walter (2024) found that many students lack the necessary GenAI knowledge to use these tools effectively. Although students generally agreed that GenAI is useful and easy to use, their self-efficacy remained low. The subtracted ENA graph showed that there is no connection between GenAI usage and high self-efficacy, most likely because they were confident in their English proficiency. Despite this, GenAI tools have the potential to scaffold learning by providing relevant language input and feedback (Cong-Lem & Daneshtar, 2024; Vygotsky, 1978).

Interestingly, the finding that GenAI use does not significantly impact self-efficacy or predicted scores is rare and might be unique to this study. Given that GenAI is still emerging in education and its use is restricted in China, students may not yet know how to fully utilize these tools for exam preparation. Walter (2024) emphasizes the need for training to help students integrate GenAI effectively and responsibly. If used improperly, GenAI use might even decrease students' academic abilities, such as critical thinking (Zhang et al., 2024).

We observed that students with higher overall self-efficacy and stronger intrinsic and extrinsic motivation for learning English predicted their scores better. This finding aligns with Meera and Jumana's (2015) study, which suggested the relationship between self-efficacy and academic performance. Similarly, Cao (2020) found self-efficacy's influences academic achievements both directly and indirectly through cognition, emotion, and learning strategies, as well as through motivation regulation. Our qualitative results help explain this phenomenon. The subtracted ENA graph displays edge links from high self-efficacy to intrinsic motivation. Students who feel confident and study for personal satisfaction tend to judge their performance well. The Usage node, meanwhile, shows a connection to "TAM_Neg". Even GenAI users encounter functionality issues that can reduce usage. Finally, extrinsic motivation lacks a direct edge to high self-efficacy, which matches the survey finding that external incentives alone do not guarantee higher self-efficacy. Saeed and Zyngier (2012) found that intrinsic motivation is closely tied to authentic student engagement, whereas extrinsic motivation can also promote ritual engagement. This suggests that students who are motivated to learn English, regardless of the source of their motivation, feel more assured about their potential test performance. Contrary to the hypothesis and past articles, the frequency of GenAI use did not significantly influence students' predicted test scores.

Implications

Although this study found that younger students who perceived GenAI as helpful and easy to use will most likely increase their frequency in use – a conclusion that reinforces the TAM. For EFL students, the frequency of GenAI is not directly related to their predicted learning outcomes but their motivation and agency to learn a language. GenAI or technology is not as effective as we assume; most past studies which concluded that GenAI tools help students to improve their skills are conducted in controlled settings with an experimental design, and those test results might not correspond well to real-world performance (Chaytor & Schmitter-Edgecombe, 2003). In real-world contexts, technology might not even be the first choice for EFL students to improve their language skills, and the roles of GenAI in this study were merely auxiliary. This is not to deny the effectiveness of GenAI in the study nor discourage others from using it, but is a call to the educators that instead of looking at superficial methods to increase students' learning outcomes, once again, the discussion must return to why a learner learns and what motivates them and gives them agency to learn. In addition, participants' physical location also limited access to mainstream GenAI tools, hindering them from using the tool to its full potential. For them, the existing GenAI might not support English input and output.

Educators in higher education should integrate GenAI tools in ways that complement traditional learning methods rather than relying on them as primary tools for skill enhancement. GenAI tools should be used to answer more rudimentary and repetitive questions under the supervision of the course teacher to ensure the answer is correct and comprehensive in order to achieve the ideal scaffolding effect. To realize this, institutional-level training programs should be implemented to help students and faculty effectively incorporate GenAI into their study and teaching routines, enhancing their ability to use these tools to overcome various learning and training challenges. In addition, when incorporating any GenAI or technologies into the classroom, there should be an adequate introduction of its usage and make sure its use is within policy and ethical guidelines.

Educators should consider polling before each class and pay more attention to older students who have relatively lower previous grades and learn with extrinsic motivation – these are the students who might require more help in technology. Educators should consider additional support when implementing GenAI in their classrooms.

Given the situation uncovered in this article, for both research institutes and GenAI developers, it is crucial to explore better ways to incorporate existing and accessible GenAI tools into learning and working processes, create or fine-tune GenAI tools that are suitable for specific needs, and allow free or affordable and equitable access to these technologies by providing the necessary devices, reliable internet access, and technical support to all students. From a policy perspective, it is essential for policymakers to support and give opportunities for a balanced approach that integrates technology with traditional teaching methods. Policies should emphasize the importance of fostering intrinsic motivation and learners' agency through professional development programs for educators focusing on these areas. Lastly, unique to our participants, the policy should also remove restrictions to certain websites that allow students to access GenAI tools.

Limitations and Future Studies

Certain aspects of this study could be improved on in future research. First, the recruitment of participants primarily from the population of Chinese students may limit the generalizability of findings to other cultural contexts. Additionally, restrictions on GenAI access, such as ChatGPT, mean that participants' experiences may not fully reflect global trends in GenAI usage for language learning. The study also relied on predicted scores rather than actual exam results, which may impact the reliability

of the findings related to academic outcomes. Furthermore, the lack of control over the types of GenAI tools used by participants could obscure differences in effectiveness across various platforms. Lastly, by including participants who are still preparing for the CET-4 test at junior or senior grade, who had prepared for longer and had more exam experience, the findings may not generalize to those who are preparing for their first CET-4 exam. These limitations suggest that future research should include cross-cultural studies, investigations using actual exam scores, and detailed analyses of specific GenAI tools in diverse language learning contexts. Despite these constraints, the study provides a foundation for understanding the relationships among GenAI use, language learning outcomes, self-efficacy, and motivation in the Chinese educational context. We would make a similar suggestion to that of Tian et al. (2024) in their article, namely that future studies should further explore the adoption of GenAI in academic context (Tian et al., 2024). Future studies should also explore the long-term effects of GenAI use, cross-cultural impacts, and the specific features of GenAI tools that contribute most to students' learning. Additionally, examining how personality influences the adoption of GenAI could provide valuable insights, enabling educational policies to better leverage GenAI technologies to enhance learning outcomes.

Conclusion

This study examined the impact of GenAI tools on EFL learners, focusing on self-efficacy and motivation toward standardized exams in China. Our findings suggest that although GenAI tools are perceived as useful and easy to use, their frequency does not significantly increase their self-efficacy. Perceived usefulness and ease of use can increase usage frequency but do not directly correlate with improved predicted learning outcomes, which reinforced the TAM model. Despite the positive perceptions of GenAI tools, the study did not find a significant enhancement of users' self-efficacy. This result suggests the need for comprehensive training and integration strategies to maximize the benefits of GenAI tools in real-world educational settings. Future research should address these limitations by conducting cross-cultural studies and utilizing actual exam scores to better understand the impact of GenAI tools on learning outcomes. Additionally, detailed analyses of specific GenAI tools in various language learning contexts are recommended to provide more insights.

Declarations

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors claim no conflict of interests.

References

- Ammari, T., Chen, M., Zaman, S. M. M., & Garimella, K. (2025). *How students (really) use ChatGPT: Uncovering experiences among undergraduate students*. arXiv. <https://doi.org/10.48550/arXiv.2505.24126>

- Andrade Preciado, J. S. A., & González Vallejo, R. (2025). Self-efficacy and perception in the use of GAI tools in the translation classroom. *Seminars in Medical Writing and Education*, 4, 526. <https://doi.org/10.56294/mw2025526>
- Anjum, F., Raheem, B. R., & Ghafar, Z. N. (2024). The impact of ChatGPT on enhancing students' motivation and learning engagement in second language acquisition: Insights from students. *Journal of E-learning Research*, 3(2), 1–11.
- Annamalai, N., Bervell, B., Mireku, D. O., & Andoh, R. P. K. (2025). Artificial intelligence in higher education: Modelling students' motivation for continuous use of ChatGPT based on a modified self-determination theory. *Computers and Education: Artificial Intelligence*, 8, 100346.
- Ardasheva, Y., Tong, S. S., & Tretter, T. R. (2012). Validating the English Language Learner Motivation Scale (ELLMS): Pre-college to measure language learning motivational orientations among young ELLs. *Learning and Individual Differences*, 22(4), 473–483.
- Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology*, 31(3), 597–606.
- Balci, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies*, 3(1), 66–82. <https://www.ijces.net/index.php/ijces/article/view/107>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age Publishing.
- Cao, Y. (2020). *The role of self-efficacy in improving English performance* [Conference paper]. Proceedings of the 2020 International Conference on Education, Social Sciences and Humanities. http://proceedings-online.com/proceedings_series/SH-EDUCATION/ESES2020/ices01080.pdf
- Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, 8(1), 1. <https://doi.org/10.29140/ajal.v8n1.102418>
- Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments*, 10(1), 60. <https://doi.org/10.1186/s40561-023-00269-3>
- Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13(4), 181–197. <https://doi.org/10.1023/B:NERV.0000009483.91468.fb>
- Cong-Lem, N., & Daneshfar, S. (2024). Generative AI and second/foreign language education from Vygotsky's cultural-historical perspective. In H. P. Bui & E. Namaziandost (Eds.), *Innovations in technologies for language teaching and learning*, Vol. 1159 (pp. 175–188). Springer. https://doi.org/10.1007/978-3-031-63447-5_10
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-methods research: A discussion on its types, challenges, and criticisms. *Journal of Practical Studies in Education*, 2(2), 25–36.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation,

- processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. E. Arnold.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House.
- Hernández-Rivera, E., Kalogeris, A., Tiv, M., & Titone, D. (2024). Self-evaluations and the language of the beholder: objective performance and language solidarity predict L2 and L1 self-evaluations in bilingual adults. *Cognitive Research: Principles and Implications*, 9(1), 75.
- Hou, Y. (2024). Effects of student choice on intrinsic motivation and language proficiency in high school L2 education. *Communications in Humanities Research*, 46, 117–122. <https://doi.org/10.54254/2753-7064/46/20242276>
- Huang, J. (2024). Enhancing EFL Speaking feedback with ChatGPT's voice prompts. *International Journal of TESOL Studies*, 6(3), 4–13.
- Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics*, 1, 102324–102324.
- Humlum, A., & Vestergaard, E. (2025). The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences*, 122(1), e2414972121. <https://doi.org/10.1073/pnas.2414972121>
- Jadhav, S., Vhatkar, S., & Aalam, Z. (2024). Bridging the gap: Exploring the revolutionary application of GenAI in language teaching and learning. *Journal of Electrical Systems*, 20(4), 2185–2193.
- Jang, J. (2022). An exploratory study on learner agency and second language writing practices of Korean high school students. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), 1–15. <https://doi.org/10.1186/s40862-022-00158-1>
- Johnson, D. M., Doss, W., & Estep, C. M. (2024). Using ChatGPT with novice Arduino programmers: Effects on performance, interest, self-efficacy, and programming ability. *Journal of Research in Technical Careers*, 8(1), 1.
- Kittredge, A. K., Hopman, E. W., Reuveni, B., Dionne, D., Freeman, C., & Jiang, X. (2025). Mobile language app learners' self-efficacy increases after using generative AI. *Frontiers in Education*, 10, 1499497. <https://www.frontiersin.org/journals/education/articles/10.3389/educ.2025.1499497/full>
- Kurtz, G., Amzalag, M., Shaked, N., Zaguri, Y., Kohen-Vacs, D., Gal, E., Zailer, G., & Barak-Medina, E. (2024). Strategies for integrating generative AI into higher education: Navigating challenges and leveraging opportunities. *Education Sciences*, 14(5), 1–20. <https://doi.org/10.3390/educsci14050503>
- Lee, Y. J., & Davis, R. O. (2024). A case study of implementing generative AI in university's general English courses. *Contemporary Educational Technology*, 16(4), ep533. <https://eric.ed.gov/?id=EJ1449352>
- Leong, J., Pataranutaporn, P., Danry, V., Perteneder, F., Mao, Y., & Maes, P. (2024). Putting things into context: Generative AI-enabled context personalization for vocabulary learning improves learning motivation [Conference session]. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3613904.3642393>
- Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications*, 11(1), 1268. <https://doi.org/10.1057/s41599-024-03755-2>
- Li, Z., Wang, C., & Bonk, C. J. (2024). Exploring the utility of ChatGPT for self-directed online language learning. *Online Learning*, 28(3), 157–180.
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, 1–21. <https://doi.org/10.1016/j.compedu.2023.104977>

- Liu, W. C., Wang, J. C. K., & Ryan, R. M. (Eds.). (2016). *Building autonomous learners: Perspectives from research and practice using self-determination theory*. Springer.
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11(1), 50. <https://doi.org/10.1186/s40561-024-00342-5>
- MacNealy, M. S. (1999). *Strategies for empirical research in writing*. Allyn and Bacon.
- Mahande, R. D., Fakhri, M. M., Suwahyu, I., & Sulaiman, D. R. A. (2025). Unveiling the impact of ChatGPT: Investigating self-efficacy, anxiety and motivation on student performance in blended learning environments. *Journal of Applied Research in Higher Education*, in press. <https://doi.org/10.1108/JARHE-07-2024-0372>
- Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in Quantitative Methods for Psychology*, 3(2), 51–59.
- Meera, K. P., & Jumana, M. K. (2015). Self-efficacy and academic performance in English. *Research in Pedagogy*, 5(2), 25–30.
- Miller, R. (2024). Pandora's can of worms: A year of generative AI in higher education. *Portal: Libraries and the Academy*, 24(1), 1–5. <https://doi.org/10.1353/pla.2024.a916988>
- Muniandy, J., & Selvanathan, M. (2025). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a public university, Malaysia. *Teaching Public Administration*, 43(1), 4–20. <https://doi.org/10.1177/01447394241230152>
- Nielsen, T., Makransky, G., Vang, M. L., & Dammeyer, J. (2017). How specific is specific self-efficacy? A construct validity study using Rasch measurement models. *Studies in Educational Evaluation*, 53, 87–97.
- Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50(1), 57–85. <https://doi.org/10.1111/0023-8333.00111>
- Pan, Z. (2019). Struggling between national pride and personal empowerment: The language ideologies held by Chinese university students towards China English. *Lingua*, 227, 102702. <https://doi.org/10.1016/j.lingua.2019.06.003>
- Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.
- Ruis, A., Siebert-Evenstone, A., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). *Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking*. Retrieved 28 August 2025 from <https://repository.isls.org/handle/1/4395>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Saeed, S., & Zyngier, D. (2012). How motivation influences student engagement: A qualitative case study. *Journal of Education and learning*, 1(2), 252–267.
- Sardi, J., Candra, O., Yuliana, D. F., Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy*, 15(1), 94–108.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.
- Shorten, A., & Smith, J. (2017). Mixed methods research: Expanding the evidence base. *Evidence-Based Nursing*, 20(3), 74–75. Royal College of Nursing. <https://ebn.bmj.com/content/20/3/74/>
- Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2017). In search of conversational grain size: Modeling semantic structure using moving stanza windows. *Journal of Learning Analytics*, 4(3), 123–139.
- Song, D. (2024). How learners' trust changes in generative AI over a semester of undergraduate courses. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-024-00446-6>

- Sun, Y., & Gao, F. (2020). An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research and Development*, 68(3), 1181–1198. <https://doi.org/10.1007/s11423-019-09733-9>
- Tian, W., Ge, J., Zhao, Y., & Zheng, X. (2024). AI chatbots in Chinese higher education: Adoption, perception, and influence among graduate students—An integrated analysis utilizing UTAUT and ECM models. *Frontiers in Psychology*, 15, 1–15. <https://doi.org/10.3389/fpsyg.2024.1268549>
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246.
- UNESCO. (2024). *UNESCO survey: Less than 10% of schools and universities have formal*. Retrieved 28 August, 2025 from <https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai>
- Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, 21(3), 323–349.
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 1–15. <https://doi.org/10.1186/s41239-024-00448-3>
- Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, 29(18), 24963–24990. <https://doi.org/10.1007/s10639-024-12819-4>
- Xie, Y., Xie, Z., Chen, S., Shen, L., & Duan, Z. (2025). An analysis of predictive modeling and factors influencing CET-4 pass rate among Chinese university students: A machine learning-based approach. *Education and Information Technologies*, 30(3), 3669–3690. <https://doi.org/10.1007/s10639-024-12964-w>
- Xu, S., Chen, P., & Zhang, G. (2024). Exploring the impact of the use of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment. *Heliyon*, 10(21), e39845. <https://doi.org/10.1016/j.heliyon.2024.e39845>
- Yan, J., & Huizhong, Y. (2006). The English proficiency of college and university students in China: As reflected in the CET. *Language, Culture and Curriculum*, 19(1), 21–36. <https://doi.org/10.1080/07908310608668752>
- Yildiz, C. (2024). ChatGPT Integration in EFL education: A path to enhanced speaking self-efficacy. *Novitas-ROYAL (Research on Youth and Language)*, 18(2), 167–182.
- Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4, 100–147.
- Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, 21(1), 34. <https://doi.org/10.1186/s41239-024-00467-0>
- Zheng, Y., Wang, Y., Liu, K., & Jiang, M. (2024). Examining the moderating effect of motivation on technology acceptance of generative AI for English as a foreign language learning. *Education And Information Technologies*, 29, 23547–23575. <https://doi.org/10.1007/s10639-024-12763-3>

Appendix A

Interview Questions and Related Research Questions (RQs)

#	Interview question	RQs
1	What year are you in university, and what is your current major?	–
2	Is your motivation for learning English related to your attitude towards AI? Why or why not?	RQ1
3	Can you describe whether your intrinsic motivation for English helps you feel more confident and capable of achieving good results in the exam? How does this intrinsic motivation affect your exam preparation and experience?	RQ1
4	Do you use AI in your daily life and studies? Have you encountered any difficulties when using AI?	RQ1
5	Do you think your attitude towards AI affects your ability to complete various parts of the CET-4 exam? Why or why not?	RQ2
6	Does the frequency of your AI usage affect your ability to complete various parts of the CET-4 exam? Why or why not?	RQ2
7	Does the frequency of AI usage affect your prediction of your final CET-4 score? Why or why not?	RQ2
8	Which part of the CET-4 exam do you think is the most important? If you feel you did well on this part during the exam, would it affect your prediction of your final CET-4 score?	RQ2

Appendix B

Qualitative Data Codebook

Table 1 *Codebook*

Code	Definition
Usage	Participant reports using GenAI more than few times a week
Usage_None	Participant states they never or almost never use GenAI.
Usage_Academic	GenAI is applied to study-related tasks such as coursework, essay writing, and programing
Usage_DailyLife	GenAI is used as interlocutor or quick fact-checking medium.
Motiv_Intrinsic	Learns English out of personal interest, enjoyment, or cultural curiosity.
Motiv_Extrinsic	Learns English for grades, employment prospects, or competition with peers.
SelfEff_High	States strong confidence in their English ability.
SelfEff_Low	Expresses weak confidence in their English ability.
TAM_Pos	Finds GenAI useful and/or easy to operate.
TAM_Neg	Finds GenAI not useful and/or difficult to operate.
Benefit_English	Claims GenAI helps to build vocabulary, grammar, fluency, or comprehension.
Difficulty_Barriers	Any obstacle to GenAI use (cost, access limits, ads, prompt input, outdated version, or school restrictions).
Prep_Facilitated	Interest or GenAI use makes preparation easier, faster, or less stressful.
Prep_Resistance	Low interest or other factors make preparation difficult.
PredInfluence_Yes	Participant says AI usage frequency or a specific exam section has changed their expected CET-4 score.
PredInfluence_No	Participant says these factors has not changed their expected score.

List of Abbreviations

Abbreviation	Full name
GenAI	Generative Artificial Intelligence
EFL	English as a Foreign Language
CET-4	College English Test 4
TAM	Technology Acceptance Model
LST	Listening Test
RST	Reading Test
WRT	Writing Test
TRS	Translation Test
EXM	Overall Exam Skills
ENG_IN	Intrinsic Motivation
ENG_EX	Extrinsic Motivation
AIU	Perceived Usefulness of GenAI
AIE	Perceived Ease of Use of GenAI
BHI	Behavioral Intention
ATT	Attitude
AIF	Frequency of Use
Prvs_Eng	Previous Score
ENA	Epistemic Network Analysis