Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7, 47-66
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

Patterns of generative AI Use in Chinese university Spanish education: A two-sample descriptive study of student and teacher perspectives

Chenyi Liao[1], Hongchun Du[2], Jian Yao[3*]
[1]Yunnan Institute of Industry Education Integration Science, China; 2895330389@qq.com (C.L.).
[2]Chinese International College, Dhurakij Pundit University, Thailand; 2690156722d@gmail.com (H.C.).
[3]LiaoNing Finance and Trade College, China; Cookieann99@163.com (J.Y.).

Abstract: Generative AI (GenAI) has proliferated in language education, yet empirical research on its real-world use remains scarce, especially beyond mainstream EFL settings. This parallel cross-sectional survey, partnered with an online research team, recruited 404 college Spanish as a foreign language (SFL) students and 406 SFL teachers across 17–18 Chinese provinces. Two custom questionnaires, built on cognitive offloading theory and critical AI application studies, measured GenAI usage, cognitive attitudes, participation tendencies, and teaching practices. Results show prevalent multi-GenAI adoption among students (only 14.1% rarely use such tools), with ChatGPT, DeepSeek, Kimi, Gemini, and Doubao mainly used for grammar, translation, and vocabulary learning. Teachers widely integrate GenAI into instruction but lack formal training: 44.6% learned it independently, and 9.9% had zero formal training. Provincial data shows a negative correlation between student self-reported critical engagement and teacher-perceived student critical engagement ($r = -.53$). Confirmatory and exploratory analyses found four theoretical subconstructs (cognitive offloading, critical engagement, beliefs, learning outcomes) converged into one dimension, reflecting severe acquiescence and common method bias in Chinese online survey data. This study establishes baseline GenAI adoption data for non-EFL Asian higher education and offers methodological warnings for relevant scholarship.

Keywords: *Chinese higher education, Cognitive offloading, Common method bias, Critical engagement, Generative AI, Online panel data, Spanish as a foreign language, Teacher beliefs.*

1. Introduction

In the past three years, the application of artificial intelligence in the field of language education has undergone profound and rapid changes: its development trajectory has rapidly evolved from relying on regular and preset teaching tools in the early days to a complex generative artificial intelligence (GenAI) system that can dynamically generate fluent and highly contextual language output according to learners' needs [1, 2]. Since the end of 2022, applications with large-scale language model (LLM) as the core technology (such as ChatGPT) have been opened to the public, the use rate of this field has increased explosively in the student community, and the attention of relevant academic research has also risen sharply - the focus of empirical research has expanded from the preliminary evaluation of the effectiveness of a single tool to a systematic discussion of the teaching significance and impact of generative AI in different language learning scenarios [3]. These fundamental developments have made the essential difference between generative artificial intelligence and the early computer-assisted language learning (CALL) technology: it is no longer just a static resource library passively used by learners, but has gradually evolved into an interactive agent that can deeply interact with learners and jointly create language content.

With the growing integration of generative AI tools into academic practice, researchers began to reveal a core paradox behind its huge educational potential. On the one hand, a large number of research

© 2026 by the authors; licensee Learning Gate
History: Received: 10 April 2026; Revised: 15 June 2026; Accepted: 19 June 2026; Published: 6 July 2026
* Correspondence: Cookieann99@163.com

results consistently show that when students are assisted by generative artificial intelligence in language learning tasks, their final output task quality, writing fluency, and overall learning satisfaction can be significantly improved [4, 5]. However, on the other hand, more and more evidence also reveals a phenomenon of cognitive unloading, that is, learners tend to systematically outsource the thinking process that originally requires cognitive efforts to external intelligent tools; this trend may potentially weaken the deep learning and understanding that language teaching is committed to cultivating [6-8]. For the students who study Spanish as a foreign language (SFL) in Chinese universities, this contradiction is particularly acute: the cultivation of communicative competence and cross-cultural language awareness essentially depends on learners' continuous and active cognitive input and meaning construction process. The performance of cognitive unloading behavior is precisely described by some scholars as a kind of "metacognitive inertia", in this state, learners may be able to efficiently complete task output, but they bypass the key links such as comprehensive analysis of language materials, self-monitoring of their own learning process, and reflective reasoning, which are the cornerstone of cultivating real language autonomy and ability [9].

Although the academic community's understanding of these potential cognitive risks is deepening, the existing research framework is still difficult to comprehensively and deeply respond to these problems in the specific context of Spanish as a foreign language teaching (SFL). At present, although the quantitative research paradigm occupies a dominant position in this field, the academic community has not yet developed a standardized evaluation tool that can effectively and on a large scale capture the unique use patterns and characteristics of generative AI in the non-English education environment [10]. It is also noteworthy that the empirical research on the group of Spanish learners in Chinese universities is almost blank: the SFL learning environment has long been ignored and marginalized by the mainstream educational AI research, and up to now, no large-scale sample research can simultaneously and systematically investigate the cognition, attitude, and practice of SFL students and teachers on this phenomenon [11, 12]. This double neglect of the multi-stakeholder perspective and quantitative research on non-native English learners is the key academic gap that this study intends to explore and try to fill.

Whereas the original protocol for this study was designed to test a series of structural hypotheses concerning antecedents and consequences of cognitive offloading, the analytical findings reported below required a substantive shift in framing. As we report in Section 5, the theoretically distinct measurement subscales collapsed empirically onto a single dimension, a measurement outcome that, taken together with the recruitment context and the homogeneity of the response distributions, precluded inferential structural modelling. The present paper consequently reframes its contribution along two lines: (a) a benchmark descriptive characterisation of GenAI use, beliefs, and pedagogical practices in a non-EFL Asian higher-education context and (b) a methodological case study for the growing literature on AI in Chinese education that increasingly relies on online research panels for participant recruitment. The paper is organised as follows: Section 2 presents the theoretical framework that motivated the study design; Section 3 reviews the relevant literature; Section 4 details the research methodology, including the recruitment context whose implications subsequently shaped our analytical strategy; Section 5 reports descriptive and comparative findings and the measurement analysis; Section 6 discusses pedagogical and methodological implications; Section 7 concludes.

2. Theoretical Framework

2.1. Cognitive Offloading Theory

Cognitive offloading refers to individuals' active use of physical or digital structures in the environment to reduce the cognitive load of tasks, thus delegating tasks that originally need internal psychological processing to external resources [13]. The theoretical roots of this concept can be traced back to the embodied theory and extended mind theory in cognitive science. It covers a wide range of daily behaviors, from traditional paper notes to modern GPS navigation, which externalize the information storage or calculation process. Risko and Gilbert [13] further distinguished two unloading

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

types in their research: one is automatic and unconscious habitual unloading, which is usually harmless; the other is deliberate unloading through strategic decision-making, that is, learners actively give up internal processing and rely on external assistance instead. This distinction has important enlightening significance in the field of education. The reason why Bjork [14] proposed "ideal difficulties" (such as memory retrieval exercises, in-depth exploration, and interleaving problem solving) can promote deep learning is precisely that its value stems from the internal processing processes that need efforts that is deliberately avoided by cognitive unloading.

In the specific situation of language learning based on general artificial intelligence (GenAI), the phenomenon of cognitive unloading has particularly significant theoretical and practical significance. The GenAI tool enables learners to systematically delegate tasks such as vocabulary retrieval, grammar analysis, discourse structure planning, and metacognitive monitoring of their own learning process to external intelligent systems. As a result, learners can obtain optimized and highly contextual text output with little personal cognitive input. Gerlich [8] confirmed the existence of this mechanism and its potential consequences through an empirical study: the study found that frequent and unreflective use of AI aids would significantly weaken learners' critical thinking ability and the depth of self-reflection. Sachdeva and Gilbert [15] also pointed out in their basic scale development research that learners will not only transfer the memory task, but also avoid the metacognitive monitoring function that usually accompanies the memory task, showing an obvious tendency to pursue the minimization of energy consumption. These studies jointly confirm that cognitive unloading is a risk factor with solid empirical evidence caused by nonreflective and overreliance on GenAI in language learning scenarios.

At the same time, cognitive unloading theory also provides a more detailed and dialectical perspective for the overall theoretical framework of this study. Iqbal et al. [16] found in a structural equation model study of pre-service teachers that in the context of collaborative learning, the relationship between the use of general AI tools and academic achievement is affected by the positive mediating effect of learners' common metacognitive ability and cognitive unloading. This shows that cognitive unloading itself is not generally harmful. In fact, whether the specific impact is positive or negative depends largely on the level of metacognitive collaborative participation accompanying the unloading process. This conditional model, that is, the same mechanism will produce different results that hinder or promote the learning effect according to whether learners actively and deeply participate in the cognitive process or passively and superficially rely on external tools, is the key theoretical basis for this study to construct the overall analysis system around the core concept of "cognitive unloading to participation continuum".

2.2. Critical Thinking and AI Integration

The relationship between the application of genai and critical thinking is one of the most controversial and urgent core issues in the field of AI education research. Critical thinking, in a broad sense, refers to the ability of individuals to conduct systematic and autonomous cognitive assessment, reflection, and independent judgment in complex situations. It is not only the core training goal of modern language teaching but also the key literacy that can be promoted or significantly weakened by the "double-edged sword" of the application of GenAI. On the one hand, advanced AI tools can generate logical and thought-provoking questions, counterarguments, and multi-dimensional evaluative prompts based on learners' input, so as to effectively help learners expand their thinking boundaries and systematically improve their analysis, synthesis, and reasoning abilities. However, on the other hand, the convenience and powerful information-generation ability of such tools may also inadvertently encourage a tendency to rely on AI-assisted information retrieval and content generation to replace the in-depth and independent critical-evaluation process. This risk is particularly prominent among learners, especially when they lack metacognitive skills to effectively identify the accuracy, bias, and logical loopholes of AI output [17, 18].

It is noteworthy that recent empirical studies have consistently shown that well-designed teaching interventions play a decisive role in alleviating and even transforming this contradiction. Hong, et al.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

[19] revealed in a quasi-experimental study on second language writing that the cognitive-unloading mechanism partially mediates the complex relationship between the use of generative artificial intelligence and the improvement of critical thinking, and the positive and negative directions of its effect are highly dependent on the specific teaching situation: when using the structured "cognitive-unloading guidance" strategy, which explicitly requires learners to actively explain and record their thinking path and decision reasons before, during, and after interaction with AI, it will produce a significant positive mediating effect and promote deep cognitive processing; on the contrary, the unstructured and laissez-faire use of AI is prone to lead to thinking inertia and a negative mediating effect. In response, Alqarni [20] recently developed and verified a set of assessment scales for teachers to systematically measure the actual effect of AI on the cultivation of students' critical thinking after it is integrated into classroom teaching, which reflects the growing consensus in the academic community that teachers' AI-critical teaching ability, that is, how they design tasks, guide reflection, and evaluate AI output, is as important as learners' own willingness and skills for critical participation. Therefore, for Chinese college second language learners and their teachers in a generally exam-oriented teaching environment, how to go beyond the use of simple technical tools to cultivate a prudent, reflective, and constructive critical AI participation ability is not only a core literacy subject for learners, but also a severe challenge for teachers' professional development.

2.3. Chinese University Learners' Foreign Language Learning Characteristics

The unique characteristics of Chinese college students as foreign language learners constitute the third theoretical pillar, which not only profoundly affects the specific manifestations of the cognitive unloading phenomenon but also fundamentally restricts the practical space and possibility of the application of critical AI. The existing literature has made an in-depth analysis of the characteristics of these Chinese college foreign language learners and summed up some common features closely related to AI-assisted learning: first, they are obviously inclined to adopt learning strategies based on memory and compensation, such as relying on repeated recitation and mechanical exercises to master language knowledge, while the use of metacognitive strategies (such as self-monitoring and planning) and social strategies (such as collaboration and interaction) shows relative inconsistency and weakness [21, 22]. This distinctive strategic feature is likely to make Chinese foreign language learners more naturally tend to use AI tools in a passive and output-oriented way. That is, AI is mainly regarded as a compensatory tool to make up for insufficient vocabulary or grammatical errors, which is used to quickly complete homework or tests, rather than as a cognitive partner or learning scaffold that can promote in-depth language exploration and stimulate critical thinking.

Secondly, the limited contact learning environment is a significant and widespread structural feature in China's foreign language education system. Most Chinese learners are almost exclusively exposed to the target language in a formal and structured classroom environment, and seriously lack the incidental learning and communication practice unique to second language acquisition in an immersive environment and realized through real social interaction and daily life contact [23]. Especially for the learners of Chinese as a foreign language, their opportunities to contact Spanish and other non-English target language environments outside formal teaching are more scarce than those of learners of English as a foreign language; this lack of continuous language contact not only leads to their higher demand and dependence on the supplementary and situational language input provided by AI, but also makes it easier for them to accept the output generated by AI without screening and without sufficient real corpus and independent judgment reference standards, thus weakening their ability of critical assessment.

Third, the self-efficacy of reading and writing is a key regulatory factor affecting the use and degree of AI. The research of Zhou and Zhang [24] further reveals that in the specific learning situation of Chinese as a foreign language, there is a dynamic track of mutual promotion and common development between learners' self-efficacy and their ability to use strategies. Specifically, learners with low self-efficacy tend to accept and follow the suggestions provided by AI without reflection and regard them as

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

authoritative answers because of their lack of confidence in their own language ability; learners with higher self-efficacy, by virtue of stronger self-confidence and judgment, can more actively and critically examine, screen, and even question the suggestions of AI, and integrate them into their own cognitive framework, to achieve more effective learning.

3. Literature Review

3.1. Cognitive Offloading in GenAI-Assisted Language Learning

Empirical documentation of cognitive offloading in GenAI-mediated learning contexts has accelerated substantially since 2023, converging across experimental, quasi-experimental, and survey-based paradigms. Georgiou [25] reported significantly lower cognitive engagement scores in a ChatGPT condition compared to controls, characterising the mechanism as AI-induced cognitive offloading and calling for pedagogical strategies to promote active, reflective engagement with AI output. Çobanoğulları [6] similarly identified learners' tendency to substitute AI-generated grammatical explanations for their own meaning-making, particularly in tasks where AI output was perceived as authoritative.

Quantitative research based on surveys has begun to systematically reveal the structural causes and influence mechanisms behind large-scale cognitive unloading. For example, Iqbal, et al. [16] used the improved utaut2 scale to investigate the pre service teacher group. The analysis results showed that cognitive unloading and shared metacognition jointly mediated the complex relationship between the use of generative AI and academic achievement. Further research reveals that the relative weight of these two mediators is not fixed, but significantly depends on the specific context of collaborative participation. At the same time, Gerlich [8] made an in-depth analysis of a broader sample of social groups by using a critically validated critical-thinking assessment tool. The study found that even after the effective control of demographic variables, the frequent use of artificial intelligence independently predicted that individuals scored low on the two indicators of critical-thinking ability and self-reflection participation. This result provides key, relevant supporting evidence from large-scale empirical studies for the "cognitive unloading hypothesis".

However, there is still a lack of research on the specific educational background of SFL (functional language learning). Nevertheless, an exploratory study conducted by Huang and Cassany [26] provides us with valuable basic process data. Through a detailed qualitative analysis of 370 ChatGPT tips generated by ten Chinese university SFL learners in a week, they identified three main modes of interaction: lexical scaffolding, structural mediation and generative delegation. Among them, generative delegation, which is the most consistent with cognitive unloading, is most common among learners with the highest level of self-reported language anxiety. This finding strongly shows that the interaction of motivation and emotion variables with traditional cognitive variables has a profound impact on learners' dependence on AI. It is on this key insight that a solid theoretical basis is provided for the design of the assessment tool for this study, which aims to measure the quality of learners' participation and the actual frequency of use simultaneously and comprehensively.

3.2. Critical Thinking and Responsible AI Integration in Language Education

Literature research on the integration of critical thinking and artificial intelligence in language education is rapidly emerging. A recurring and thought-provoking theme is that there is a significant gap between the application speed of AI, especially generative AI, and the critical AI literacy that educators and learners need to develop. Many studies have consistently shown that language learners have widely used generative AI (genai) tools. However, their participation and mastery of the assessment ability, critical questioning, and metacognitive skills necessary for the responsible and discriminative use of these tools are quite limited [10]. In response to this core contradiction, Vendrell and Johnston [27] integrated the design principles in the field of cognitive science and learning science, and clearly pointed out that to achieve truly effective critical AI integration, educational design must actively incorporate clear metacognitive teaching strategies, embed structured reflective tasks, and take

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

systematic AI literacy training as an organic part of the curriculum system, rather than an accidental or incidental product.

In this integration process, teachers' cognition, beliefs, and teaching practice are increasingly regarded as the key mediating factors affecting learners' in-depth and critical participation in AI. The research of Toncelli and Kostka [28] recorded in detail the complex and mixed emotional reactions of language teachers to generative artificial intelligence (GenAI) - they are excited about the possibilities brought by technology, but also feel frustrated and uneasy about the possible impact on traditional teaching roles and methods, and stressed the need for institutional level to provide system support to build a teaching-practice community integrating artificial intelligence that can safely explore, collaborate and share. The findings of Zaimoğlu and Dağtaş [29] further confirm the importance of the transformation of teachers' roles. Their research shows that English teachers who successfully transform their roles from simple knowledge imparters to high-quality and critical interactors between AI and human learners tend to show higher efficiency and teaching wisdom when using AI to assist teaching. In order to evaluate this key dimension more scientifically, Alqarni [20] specially developed the "AI Critical Teaching Scale" for language teachers, which provides a reliable tool based on psychometrics to measure the extent to which teachers integrate critical thinking in teaching.

However, although the relevant research has made progress in the context of English education, the research on Spanish teachers' cognition, attitude, and classroom practice of generative AI is almost completely missing in the existing international literature. This significant blank area is the core goal of this study, which is to be directly and deeply filled.

3.3. Chinese University Learners' GenAI Engagement Patterns

Beyond the theoretical characteristics outlined in Section 2.3, a growing empirical literature addresses the specific GenAI engagement patterns of Chinese university foreign language learners. Survey research on Chinese EFL students' ChatGPT adoption has found that the tool is treated as an everyday utility rather than a specialist resource, a finding that shifts research focus from adoption decisions to engagement quality [30, 31]. At present, the academic community has gradually confirmed that cognitive and metacognitive factors are the main regulatory factors that affect the efficient use of generative AI by this group, replacing the focus on tool perception variables in early studies [32].

Yu et al. [33], in a quasi-experimental study of 100 international students at a Chinese university, found substantial variation in the quality and depth of learner engagement with ChatGPT in multilingual learning contexts, with some students exploiting its interactive potential for cultural and discourse learning while others used it primarily as a model-output generator. Li and Zhang [34], in a survey of 712 Chinese students studying abroad, found that ChatGPT use positively affected foreign language self-efficacy and enjoyment, with enjoyment mediating the self-efficacy effect, suggesting a motivational pathway through which AI use may build rather than erode learner agency under certain conditions. However, as Huang and Cassany [26] document for Chinese SFL learners specifically, high linguistic anxiety is associated with generative delegation, the most offloading-intensive AI use pattern, suggesting the anxiety-AI relationship is non-linear and potentially bidirectional.

3.4. Teachers' Perceptions of GenAI in Foreign Language Education

The research system on teachers' cognition, attitude, and practice of generative artificial intelligence (GenAI) has developed in parallel with the relevant learner-centered literature and is directly related to the teacher survey in this study. Specifically, Cabellos et al. [35] conducted a systematic survey on 321 college teachers. The results showed that teachers' cognition of the impact of GenAI education showed a distinct two-factor structure, that is, they generally regarded it as a potential opportunity or a real threat. Further analysis shows that this cognitive tendency is not generated out of thin air but is significantly affected by multiple factors such as teachers' previous experience in using AI, the degree of support provided by their institutions, and the differences in their subject areas. For example, teachers who lack experience in the use of AI tend to view GenAI as a potential threat to the cultivation of

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

critical thinking and academic integrity norms; in contrast, experienced teachers are more likely to emphasize their unique advantages in realizing personalized teaching and providing instant feedback. In addition, the behavioral intention research based on the extended technology acceptance model framework further reveals that perceived practicability and ease of use can indeed effectively predict teachers' willingness to adopt the GenAI tool. However, ethical concerns and concerns about the reliability of AI output results constitute significant obstacles, especially among language teachers whose professional identity is closely related to the quality of language output and text authenticity [28, 29].

Another important development directly related to the focus of this study is the research on genai literacy as a specific professional ability. In this direction, Wang et al. [36] developed and strictly verified a 32-item English as a foreign language teacher's GenAI Literacy Scale. Through exploratory and confirmatory factor analysis, the study finally determined the five core dimensions of literacy, namely: general knowledge, application, evaluation, design, and ethics. This empirically tested dimension structure provides a solid theoretical basis and structural framework for the design of the teachers' questionnaire in this study. Based on the scale and the specific situation and actual needs of English as a foreign language teaching, this study carefully evaluates and integrates the original items, and adds several practical items suitable for the specific teaching situation.

3.5. Research Gaps

The above review reveals three interrelated and urgent research gaps, which together constitute the core starting point of this study. First of all, in the field of artificial intelligence in language teaching, the teaching scenario of English as a foreign language (EFL) has been dominant for a long time. This imbalance has led to the fact that the learning environment of Chinese as a mother tongue (SFL), including its unique linguistic typological characteristics, relatively limited real context contact conditions, and specific learner group characteristics, has not been fully and deeply explored. Secondly, in view of the specific situation of the use of generative artificial intelligence (genai) by foreign language learners in Chinese universities, there is still a lack of systematic multi-dimensional measurement research. Most of the existing assessment tools are only verified in EFL scenarios or non-Chinese learners' samples, and their applicability and effectiveness still need to be tested in China's local context. Third, teachers' views and attitudes towards the integration and application of generative AI in SFL courses have hardly been brought into the perspective of empirical research, which leads to a significant cognitive gap in the overall teaching ecosystem formed and relied on by learners' AI use patterns. In view of this, this study aims to systematically respond to and fill in the above research gap by using a dual-perspective questionnaire designed in parallel for students and teachers.

4. Research Methodology

4.1. Research Design

This study adopts a quantitative cross-sectional survey design. The subjects of the study include two parallel sample groups, namely students majoring in Spanish as a foreign language (SFL) learning in Chinese universities and teachers engaged in Spanish teaching in Chinese universities. Before the implementation of the study, we formulated a complete original research plan in advance, which clearly planned to use confirmatory factor analysis (CFA) and structural equation modeling (SEM) as the core statistical methods, aiming to test the hypothesis and make causal inference on the theoretical relationship between cognitive unloading, critical participation, AI literacy, and learning outcomes. However, as the results of the data reported in section 5.6 and the discussion in section 6.3 show, the actual data collected did not fully support the statistical discriminant validity between the previously preset potential constructs. Therefore, according to the actual characteristics of the data, we adjusted the analysis strategy accordingly, and focused on the descriptive and comparative analysis of the actually obtained measurement structure. We hereby publicly and in detail disclose the adjustment

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

process and reasons of this analysis scheme, because the flexible and transparent report on this method exactly constitutes one of the most important and substantive academic contributions of this study.

4.2. Research Questions

After the reconstruction of the analysis framework, this study focuses on four research issues:

Research question 1: What are the usage patterns of students majoring in Spanish as a Foreign Language (SFL) in Chinese universities in using generative AI tools, including frequency of use, tool diversity, and task distribution?

Research question 2: What are the characteristics of Chinese college Spanish teachers' general cognition of the integration of generative AI into teaching, their observed student-use behavior, and their self-reported personal teaching practice?

Research question 3: How consistent or different are the results of students' self-reports and teachers' observations on the topics related to generative AI in SFL teaching scenarios?

Research question 4: Is there a regular association at the provincial level that can link the use characteristics of students' generative AI with the relevant use characteristics of teachers?

4.3. Participants and Recruitment

Both student samples and teacher samples were recruited through the commercial online research platform "Wen juanxing" (also often referred to as Sojump), which is widely used in the field of education research in China. To ensure the high correlation between the research sample and the research topic, we set clear recruitment criteria, including the following three points: (a) participants must be students who are studying Spanish as a foreign language (SFL) in Chinese universities, or in-service teachers who work as Spanish teaching in Chinese universities; (b) all participants must have at least one semester (about four months) experience in using generative AI tools in teaching activities or language learning tasks; (c) for student participants, their self-rated Spanish language ability is also required to reach the A2 level or above of the Common European Languages (CEFR). In order to encourage participation and recognize their time, researchers provided appropriate economic rewards for eligible participants through the standard incentive system of the platform. All data collection work was concentrated in May 2026 and lasted for two weeks.

Finally, this study collected a total of 810 valid samples, including 404 students and 406 teachers. The two groups of samples were basically the same size. In terms of geographical distribution, both groups of samples cover 17 to 18 provinces in mainland China, showing good geographical representation. Specifically, participants in Jiangsu province accounted for the highest proportion in both groups (33.7% in the student sample and 30.3% in the teacher sample); other provinces with a large number of participants and strong representation also include Liaoning, Fujian, Zhejiang, Anhui, Guangdong, Hubei, Hunan, Hebei, and Shandong. More detailed demographic characteristics of the sample, such as gender, age, education level, and other information, will be presented in Section 5.1 of the report.

4.4. Instruments

4.4.1. Student Questionnaire

The student questionnaire used in this study consists of five main parts. Part A aims to collect students' background information, specifically covering gender, current grade, years of Spanish learning, frequency of using generative artificial intelligence (genai), AI tools preferred in various learning tasks (this topic is a multiple-choice question), and the types of tasks usually assigned by teachers or courses using AI (multiple-choice questions). Part B contains six items, which are specially used to evaluate the possible cognitive unloading behavior of students in the process of Spanish learning; these items are adapted from the research of Sachdeva and Gilbert [15] and Gerlich [8] and have been localized in the specific context of foreign language learning. There are seven items in Part C, which focus on assessing students' metacognitive awareness and critical participation when using genai.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

The item design refers to the research results of Pan et al. [30] and Ng et al. [32]. In part D, students' cognition and attitude towards the role of GenAI in language learning are investigated through eight items, including the items selected from the study of Cabellos et al. [35], which also cover the potential benefits and concerns perceived by students. Part E contains six items, which are mainly used to measure students' self-reported perceived learning outcomes. The items come from the research of Wang et al. [31] and Xingzun and Hongtao [37]; Articles 12, 17, and 30 in this part adopt the reverse scoring method to effectively reduce the possible compliance deviation in the questionnaire response. All items in the questionnaire using the Likert scale use the six-point scoring method (1 for "strongly disagree" and 6 for "strongly agree"), which aims to encourage respondents to make a clearer choice on the scale continuum rather than simply inclining to the middle value. In order to ensure the understanding consistency of cross-language subjects, the questionnaire provided a complete bilingual version of Mandarin and Spanish. The first version is translated by an assistant who is proficient in bilingualism, and then back-translated by a second independent bilingual reviewer to ensure the accuracy and equivalence of language expression.

4.4.2. Teacher Questionnaire

The teacher questionnaire is divided into five main parts: Part A mainly investigates the professional background information of teachers, including teaching years, course levels, the frequency of using GenAI in daily teaching, and whether they have participated in formal or autonomous AI-related training. Part B consists of 11 items to assess teachers' cognition and attitude towards the application of GenAI in the field of language teaching. These items are adapted from the research scale of Cabellos et al. [35] and are adjusted according to the actual situation and needs of Spanish teaching. Part C contains 8 items and uses the theoretical framework proposed by De Vries et al. [38] for reference to measure teachers' teaching tendency from the two teaching-oriented dimensions of "replication" and "construction." In part D, there are six items, which focus on the evaluation of students' cognitive unloading phenomenon that may occur in the learning process and their critical participation behavior in the task. Part E also contains 6 items, which mainly investigate the AI-critical teaching practice implemented by teachers based on the research of Alqarni [20] and Wang et al. [36]. All parts of the questionnaire were scored with the six-point Likert scale.

4.5. Data Analytic Strategy

Following the re-framing described in Section 4.1, the analytical strategy proceeded in four stages. First, data-quality screens examined response-time distributions, item-level missingness, and within-respondent response variability (straightlining indicators). Second, sample-level descriptive statistics characterised demographic distributions and GenAI-use patterns (RQ1, RQ2). Third, item-level descriptive statistics characterised attitudes and observed behaviors on each scale, with student–teacher comparisons on conceptually parallel items conducted via independent-samples t-tests with Cohen's d effect sizes (RQ3). Fourth, province-level aggregate scores were computed for the ten provinces with at least ten respondents in both samples, and cross-stakeholder correlations at the province level were estimated (RQ4).

In addition, the measurement structure of the questionnaire instruments was examined through internal-consistency reliability analysis (Cronbach's α, McDonald's ω), exploratory factor analysis (principal-axis with oblimin rotation), and confirmatory factor analysis of the original theorised structure compared against alternative competing structures including a Harman single-factor model. Convergent validity was assessed through composite reliability (CR) and average variance extracted (AVE); discriminant validity through factor correlations, the Fornell–Larcker criterion, and the Heterotrait–Monotrait ratio [39]. The results of these measurement analyses, reported in Section 5.6, indicated that the structural equation modelling and hypothesis-testing analyses originally planned were not statistically supportable, precipitating the analytical re-framing reported throughout.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

All analyses were conducted in Python 3.12 using pandas (data manipulation), SciPy (inferential tests), factor_analyzer (EFA), and semopy (CFA). Analysis code is available from the authors on request.

4.6. Ethical Considerations

The study was conducted in accordance with institutional ethical review requirements. Participation was voluntary; participants were informed of their right to withdraw at any time, and no personally identifiable information was retained. When collecting information, the recruitment team only collected the provincial geographic information corresponding to the participants' network IP addresses. These data do not involve more detailed cities, districts, or specific locations. In the whole process of processing, we strictly ensure that they are only used in the form of aggregated statistics, which is dedicated to macro provincial level trend analysis and overall insight, and will never be used to identify or relate to any specific individual.

5. Results

5.1. Sample Characteristics

A total of 404 valid samples from students were collected in this study. In terms of gender composition, female participants accounted for 47.5%, male participants accounted for 27.0%, and 25.5% of the participants were nonbinary or chose not to disclose gender information. As far as the learning stage is concerned, the sample is mainly composed of senior undergraduates, of which the fourth-grade students account for the highest proportion, reaching 33.4%; postgraduates took the second place, accounting for 24.8%; grade-two students accounted for 16.6%, grade-three students accounted for 14.6%, and grade-one students accounted for 10.6%. Most of the students have long learning experience in Spanish: 40.8% have studied for 3–4 years, and 30.4% have studied for more than 5 years; 19.8% of the students have studied for 1–2 years, while only 8.9% of the students have studied for less than one year.

For teachers, a total of 406 valid samples were collected. In terms of gender distribution, female teachers accounted for 47.3%, male teachers accounted for 35.2%, and teachers who chose not to disclose their gender accounted for 17.5%. The distribution of teaching experience is relatively balanced: the proportion of teachers with 7–10 years' experience is the highest, which is 31.5%; teachers with less than 3 years' teaching experience accounted for 23.6%; teachers with more than 10 years' experience accounted for 22.7%; 22.2% of teachers have 3–6 years of teaching experience. From the perspective of language level, the number of teachers who teach primary courses is the highest, accounting for 47.3%; 29.3% of teachers teach intermediate courses; 23.4% of teachers teach advanced courses. See Table 1 for detailed demographic distribution.

Table 1.
Sample demographic characteristics.

Sample	Variable	Most frequent	n	%
Students (n=404)	Gender	Female	192	47.5%
	Year of study	Year 4	135	33.4%
	Years of Spanish	3–4 years	165	40.8%
	AI use frequency	Often (weekly)	164	40.6%
Teachers (n=406)	Gender	Female	192	47.3%
	Teaching experience	7–10 years	128	31.5%
	Levels taught	Beginner (A1–A2)	192	47.3%
	AI use frequency	Often (weekly)	138	34%
	AI training	Formal training	185	45.6%

5.2. Patterns of Student Generative AI Use

In the student sample, the use of generative AI tools is almost everywhere, and its high-frequency applications have become a significant norm: as high as 69.1% of students said that they use such tools at least once a week, of which 40.6% of students belong to the category of "frequently used", and 28.5%

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

of students have reached the intensive level of "daily or almost daily use"; In contrast, only 14.1% of the students said they used it "rarely", and the frequency was only once a month or less. At the same time, 16.8% of the students said they used it "sometimes", that is, several times a month. It is worth emphasizing that no students reported that they had never used generative AI in Spanish learning - this result not only conforms to the participant recruitment criteria of this study, but also objectively reflects the wide penetration and universal application of this technology in the current field of higher education in China.

Further observation shows that students do not rely on a single mainstream application in actual learning, but are active in a general AI ecosystem containing a variety of tools. ChatGPT is the most frequently mentioned tool, with 69.3% of students using it. At the same time, the usage rate of four other tools exceeds 50%, which are DeepSeek (66.3%), kimi/moonshot (63.9%), Google Gemini (58.7%), and Doublao (50.2%). In addition, one third (33.4%) of the students said they also used other AI tools outside the above list, and all students reported using at least one tool, and no one said they did not use any tools at all. This widely used multi-tool parallel usage mode is in sharp contrast to the practice in early studies (these studies are usually carried out before the widespread application of Chinese native large language models) that the use of general AI is simply equivalent to the use of ChatGPT. See Table 2 for specific differences.

Table 2.
Generative AI tools used by Chinese university SFL students (n=404; multi-select).

AI tool	n	% of students
ChatGPT	280	69.3%
DeepSeek	268	66.3%
Kimi (Moonshot)	258	63.9%
Google Gemini	237	58.7%
Doubao (ByteDance)	203	50.2%
Other	135	33.4%
None	0	0%

In terms of the distribution of the use of specific tasks, the survey data showed that students most often used generative AI tools for grammar checking (71.0%), translation (69.8%), and vocabulary acquisition (64.4%). In addition, slightly more than half of the students (55.2%) said they would apply it to writing and composition tasks. In contrast, the proportion of students who use the tools in listening and speaking practice (46.5%) and obtain the interpretation of language rules (47.3%) is about half of the total sample. It can be seen that the types of tasks used frequently are obviously concentrated in the fields related to the language system, such as grammar, vocabulary, and translation; the actual frequency of using productive skills, especially writing and speaking, is relatively low. This distinctive distribution is highly consistent with the tool-use strategies and learning priorities commonly held by Chinese college foreign language learners (see Section 2.3 for correlation analysis).

5.3. Student Attitudes and Self-Reported Engagement

The comprehensive scores of the four theoretical subscales, cognitive unloading (CO), critical participation (CE), belief in general artificial intelligence (BEL), and perceived learning outcomes (LO), are summarized in Table 3 below. Under the measurement framework of the six-point scale, the comprehensive mean value of all constructs is concentrated in the narrow range of 4.30 to 4.49, which generally reflects the respondents' moderate to positive attitude toward these measurement dimensions. The standard deviation of each subscale also showed a high degree of consistency, and its value distribution was between 1.05 and 1.18, indicating that the degree of dispersion of the data was similar. Specifically, the average score of students' cognitive unloading-related items was 4.37 (standard deviation = 1.18), the average score of students' critical participation items was 4.30 (standard deviation

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

= 1.18), and the average score of students' perceived learning achievement items was 4.31 (standard deviation = 1.18).

Because the belief subscale contains parallel measurement items oriented by interests and concerns, it is further divided into two independent subdimensions for descriptive analysis. A noteworthy finding is that the score of the dimension of concern (Bel concern, whose entries involve AI as a potential means of cheating, which may lead to reduced learning motivation, long-term harm and need to be restricted) is slightly higher than that of the dimension of interest (Bel positive, whose entries involve AI as promoting effective learning practice, saving time, providing instant feedback and improving learning participation): the average score of the dimension of concern is m=4.49 (standard deviation sd=1.24), while the average score of the dimension of interest is m=4.36 (sd=1.21). This scoring trend reveals, from the descriptive level, that Chinese foreign language learners do not show a unified and unreserved enthusiasm for general AI; On the contrary, the potential benefits brought by AI and the accompanying worries within the group coexist with considerable intensity, forming a complex picture of their attitude cognition.

The descriptive data at the item level further illustrates the above overall situation. In the student data set, the item with the highest score is "I worry that relying on artificial intelligence will damage my long-term Spanish learning effect" (the average score is 4.62 points, and 64.6% of the respondents give a high score of 5 or 6 points). Followed by "AI tools can provide immediate and practical feedback for my Spanish learning" (average score 4.50, 62.0% of respondents gave 5 or 6 points), and "my Spanish level has improved since I started using AI" (average score 4.48, 59.4% of respondents gave 5 or 6 points). The trend at the item level clearly shows that, on the one hand, students highly recognize the significant practicability and auxiliary value of AI tools in language learning; on the other hand, they have strong concerns about the long-term negative impact it may bring. This recognition reflects an obvious ambivalence rather than a general, one-way optimistic attitude.

Table 3.
Student composite scores by subscale (n=404; six-point Likert; 1=Strongly Disagree, 6=Strongly Agree).

Subscale	Items	M	SD	Cronbach's α
Cognitive Offloading (CO)	6	4.37	1.18	0.840
Critical Engagement (CE)	7	4.30	1.18	0.856
Beliefs – benefit (BEL-pos)	4	4.36	1.21	0.844
Beliefs – concern (BEL-con)	4	4.49	1.24	0.797
Perceived Learning Outcomes (LO)	6	4.31	1.18	0.836

5.4. Teacher Beliefs, Observations, and Practices

Among teachers, the application of artificial intelligence has been quite popular: the survey shows that 65.5% of teachers said that they would use artificial intelligence tools at least once a week in their daily teaching practice, of which 34.0% were "frequently" used, and 31.5% even reached the frequency of "daily use." However, there is a significant gap in the corresponding training level: Although 45.6% of the teachers said they had received systematic training in the application of AI in the field of education, 44.6% of the teachers mainly relied on autonomous learning to master relevant knowledge, and 9.9% of the teachers had not received any form of training, although they had used AI tools in teaching. Combined with the data that only 1.7% of teachers said they had never used AI in teaching, this distribution trend shows that about half of the Spanish teachers surveyed by us independently explored and integrated AI technology without formal and systematic teaching training.

The comprehensive scores of teachers on the six subscales are summarized in Table 4. The average score of each subscale is between 4.30 and 4.39, and the overall score is highly centralized with little difference. Further analysis of these average values shows the following significant characteristics: (a) teachers' "reproducibility teaching orientation" (mean = 4.34) and "constructive teaching orientation" (mean = 4.38) are basically the same, indicating that teachers in this sample do not favor a single teaching idea, but are compatible with and hold the two orientations at the same time; (b) In the

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

artificial intelligence belief subscale, the score weights of "benefit perception" and "worry perception" are equivalent (the average is 4.28 and 4.30, respectively), which reflects that teachers are aware of the ambivalence of students in the face of artificial intelligence; (c) The scores of the observed teaching shirking behavior (mean = 4.39) and critical participation behavior (mean = 4.34) are similar, which shows that, in the view of teachers, these two different types of student behavior patterns are presented to a certain extent in their teaching environment.

Table 4.
Teacher composite scores by subscale (n=406).

Subscale	Items	M	SD	Cronbach's α
Beliefs about AI – benefit (TBEL-pos)	5	4.28	1.18	0.832
Beliefs about AI – concern (TBEL-con)	4	4.30	1.19	0.815
Reproductive Teaching Orientation	4	4.34	1.20	0.794
Constructive Teaching Orientation	4	4.38	1.21	0.789
Observed Student Offloading	4	4.39	1.18	0.771
Observed Student Critical Engagement	2	4.34	1.20	0.656
Teaching Practice with GenAI	6	4.37	1.18	0.846

At the level of specific projects, the views most recognized by the teachers' group are mainly concentrated in three aspects: first, they believe that "the application of artificial intelligence technology makes it more difficult for teachers to accurately assess students' real Spanish level", and the average support score for this view is as high as 4.47 points; Secondly, "in language courses, clear and operable institutional policies must be formulated to regulate the use of AI", the average score of this item is 4.35; In addition, teachers also generally agree that "the effective integration of AI into Spanish teaching practice requires teachers to invest a lot of extra preparation time", which also obtained an average score of 4.35. These highly supported views clearly focus on three key areas: teaching evaluation, policy norms, and workload management. This not only reveals the concerns of teachers about the potential risks in the teaching process but also profoundly reflects the extent to which their concerns have extended to the improvement of the support system and professional infrastructure at the level of educational institutions, indicating that the existing mechanism still has obvious deficiencies in dealing with the integration of new technologies.

5.5. Student–Teacher Comparison on Parallel Themes

This study uses seven groups of conceptually parallel assessment items to systematically test the consistency and differences between the results of students' self-reports and the beliefs or observations held by teachers (see Table 5 for specific items). Through the analysis of independent sample t-test, it is found that on most parallel topics, there is only a small and statistically insignificant difference between the average score of students and teachers (the fluctuation range of Cohen's D value is between −0.14 and+0.20). This result shows that students and teachers have a highly consistent recognition tendency on the surface measurement index of AI-related beliefs. In all pairs of evaluation items, the largest difference between teachers and students was observed in the evaluation of the function of "artificial intelligence can provide instant feedback" (the average student score M=4.50, the average teacher score M=4.28). The data showed that students' recognition of this advantage was significantly higher than that of teachers.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

Table 5.
Student–teacher comparison on conceptually parallel items.

Theme	Stu M (SD)	Tch M (SD)	Diff	Cohens d	p
Immediate feedback	4.50 (1.51)	4.28 (1.70)	+0.22	+0.14	0.06
AI reduces motivation	4.45 (1.49)	4.31 (1.53)	+0.14	+0.09	0.21
Access to useful information	4.24 (1.65)	4.14 (1.59)	+0.10	+0.06	0.40
Critical evaluation of AI	4.22 (1.50)	4.17 (1.58)	+0.05	+0.03	0.66
Pedagogical creativity/engagement	4.32 (1.57)	4.30 (1.57)	+0.02	+0.01	0.85
Concerns about over-reliance	4.62 (1.50)	4.25 (1.59)	+0.36	+0.24	0.001
Need for guidance/limits	4.46 (1.57)	4.35 (1.59)	+0.11	+0.07	0.34

Among all the subjects surveyed, the only one with statistically significant differences between teachers and students (P<0.01) was "worry about over dependence". The specific data show that the students' worry about this is significantly higher than the teachers' group, in which the average score of students is 4.62, while the average score of teachers is 4.25, and the effect size D is +0.24, indicating that this difference has practical observation significance. This finding deserves special attention and in-depth discussion, because it is contrary to the usual assumption that "students may generally lack critical thinking, while teachers are more cautious and vigilant." The actual data clearly reveals that students' concern about the potential risks of long-term overreliance on technology or external support is at least at the same level as teachers', and even more prominent in some dimensions.

5.6. Province-Level Patterns and Cross-Stakeholder Correlations

Province-level aggregation was conducted for the ten provinces with at least ten respondents in both the student and teacher samples (Anhui, Fujian, Guangdong, Hebei, Hubei, Hunan, Jiangsu, Liaoning, Shandong, Zhejiang). Cross-stakeholder correlations at the province level revealed three patterns of interest. First, student self-reported cognitive offloading was essentially uncorrelated with teacher-observed student offloading ($r = -.03$, $p = .94$), indicating no province-level concordance between learner self-reports and teacher observations on this behavior. Second, student self-reported critical engagement was negatively associated with teacher-observed student critical engagement ($r = -.53$, $p = .12$; non-significant given small province-level n but substantively large): in provinces where students reported higher critical engagement, teachers observed less of it. Third, student positive beliefs about AI showed a modest positive association with teacher positive beliefs about AI ($r = .31$, $p = .39$): some convergence on attitudinal level, but again non-significant with the small province count.

The negative association between student self-reports and teacher observations on critical engagement is the most theoretically suggestive of these patterns. It is consistent with the broader literature on metacognitive self-perception, in which less-experienced learners tend to overestimate their own analytical engagement relative to objective indicators (the Dunning–Kruger pattern). For the present study, however, given the small number of province-level observations ($n = 10$), this finding should be interpreted as hypothesis-generating rather than confirmatory.

5.7. Measurement Structure and Common Method Bias

The measurement analyses produced a coherent and consequential pattern. Internal-consistency reliability was high for all subscales (student α range: .80–.88; teacher α range: .66–.91). Confirmatory factor analyses of the originally specified four-factor (student) and six-factor (teacher) models obtained nominally acceptable fit indices: for the student model, $\chi^2(318) = 460.86$, $p < .001$, CFI = .976, TLI = .973, RMSEA = .033, SRMR = .029; for the teacher model, $\chi^2(419) = 632.98$, $p < .001$, CFI = .970, TLI = .967, RMSEA = .036, SRMR = .029.

However, three subsequent findings indicated that these acceptable fit indices were not interpretable as evidence of construct distinctness. First, a Harman single-factor confirmatory model, in which all items load on a single latent factor, fit the data nearly as well as the multi-factor models (student: CFI = .972, RMSEA = .035; teacher: CFI = .970, RMSEA = .035), and the difference in fit between the multi-

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

factor and single-factor specifications was small (ΔCFI ≤ .006). Second, exploratory factor analysis on both datasets produced a single dominant factor accounting for over 49% of total variance (student: 49.4%; teacher: 50.3%), exceeding the 50% threshold conventionally taken to indicate substantial common method bias [40]. Third, factor inter-correlations in the standardized solutions of the multi-factor models exceeded 0.99 across all factor pairs, indicating no statistical discriminant separation among the theorized constructs. The Heterotrait-Monotrait ratio (HTMT) similarly exceeded 1.0 across all factor pairs, far above the .85 threshold for discriminant validity [39]. Average variance extracted (AVE) for all subscales fell below the .50 threshold (range: .46–.49), failing the Fornell-Larcker criterion.

Taken together, these results indicate that the theoretically distinct constructs of cognitive offloading, critical engagement, beliefs about AI, and perceived learning outcomes, as well as the corresponding teacher-side constructs, collapsed onto a single underlying response dimension in both samples. This pattern is consistent with severe acquiescence response bias and common-method variance, which can produce inflated factor intercorrelations independently of the substantive content of the items [40, 41]. The patterned homogeneity of item means (item-level M range: 4.14–4.62), standard deviations (range: 1.45–1.73), and response time distributions (median 200 seconds, SD ≈ 42 seconds across both samples), combined with the absence of item-level missingness on a platform that permitted blank responses, jointly support a common-method-bias interpretation.

On the basis of this measurement evidence, the originally planned structural equation models testing antecedents and consequences of cognitive offloading were judged not statistically defensible and were not estimated. The reported results therefore concentrate on descriptive and comparative findings whose interpretation is robust to the measurement homogeneity observed.

6. Discussion

6.1. Patterns of GenAI Use in Chinese University SFL

The descriptive research results reported in sections 5.2 and 5.4 clearly establish two key empirical models of substantial significance in this field. First of all, the use of generative artificial intelligence (genai) by students in the field of language learning (SFL) in Chinese universities has shown the distinctive characteristics of generalization and diversification of tools. Specifically, more than two-thirds of students use at least one genai tool every week, while most students actually use a mixture of three to five different tools in the process of language learning. This directly challenges and overturns a leading assumption in early academic research, that is, in the Chinese context, using genai is basically equivalent to using ChatGPT in function. Current data shows that the popularity of Chinese tools developed in China (such as DeepSeek, Kimi, and Doulao) and international mainstream alternatives (such as Gemini) among students is equal to that of ChatGPT, and each tool is adopted by 50% or more students. This ecosystem composed of a variety of tools has an important impact on instructional design and future research directions: on the one hand, only focusing on the comparative study of ChatGPT tools, its conclusions may not accurately reflect the real experience and effect differences of contemporary students in the multi-tool environment, thus limiting its theoretical explanatory power and universality; on the other hand, the applicability and effectiveness of teaching interventions and resources developed around the characteristics of a single tool may also face challenges in the real situation where students actually shuttle between different tools, because students' use behavior and learning strategies are likely to become more complex and dynamic due to tool switching.

Second, the teacher data reveal a substantial training infrastructure gap. Almost half of Spanish teachers in our sample reported integrating AI into teaching without any formal training, and 9.9% reported no training of any kind. This finding extends the EFL-focused observations of Toncelli and Kostka [28] regarding teachers' affective ambivalence to a Spanish-specific context and points to an urgent institutional priority: the development of teacher professional development resources adequate to the scale and pace of GenAI adoption that is already occurring.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

6.2. Convergence and Divergence Between Stakeholder Perspectives

The student–teacher comparison reported in Section 5.5 produced two findings worth emphasising. The first is the modest extent of student–teacher disagreement on most parallel themes: on six of seven comparison points, mean differences were below 0.25 on the six-point scale, and Cohen's d values fell below 0.20. Students and teachers in this sample broadly share their assessment of GenAI's affordances and risks in language learning.

The second significant difference is that students' worry about excessive dependence on AI is significantly higher than that of teachers (effect d=+0.24, P<0.01). This discovery challenges a common assumption in public discussions and even some academic views that students tend to have an uncritical, optimistic, and receptive attitude towards AI, while teachers are generally more cautious because of their experience and responsibility. The actual data reveal a more complex picture: while actively using technology, students also show a stronger sense of self-reflection. They are more clearly aware of the potential risks and vulnerabilities that excessive reliance on artificial intelligence may bring in the process of their own cognitive development and skill formation. Of course, whether this cognitive alertness can be effectively transformed into behavioral adjustment in learning practice cannot be concluded by self-report data alone, which needs to be verified by further behavioral research.

In addition, another noteworthy phenomenon is that in the summary data at the provincial level, there is a moderate negative correlation between students' self-reported critical participation in AI and students' actual critical participation observed by teachers (r=−0.53). Although limited by the small number of samples in provinces, the estimation of this correlation coefficient has a large uncertainty range; it is undoubtedly one of the most thought-provoking directions in this study and the most worthy of future research. This finding is highly consistent with the theoretical characteristics of "metacognitive calibration bias," that is, learners tend to overestimate their level of actual analysis and critical participation relative to objective, observable indicators. This is also consistent with the existing conclusion in the field of self-regulated learning, that is, individuals' judgment of their own learning process is not always accurate. If this finding is further confirmed in future research samples of a larger scale and covering multiple types of educational institutions, it will have direct and important teaching practice enlightenment: students may not be able to independently and accurately perceive their own tendency of "task unloading" or "thinking outsourcing" in learning. Therefore, when evaluating the quality of students' AI use, teachers' systematic observation, evaluation, or introduction of external participation quality indicators can be used as an important and necessary supplementary means for students' self-evaluation tools.

6.3. Methodological Reflections: Online Panel Data and Common Method Bias

The most important finding of this study, that is, the different measurement constructs in theory in the two samples are all attributed to a single, dominant response dimension, is worthy of in-depth methodological reflection. The impact of this discovery is not limited to the study itself, but also extends to the literature system of AI research in the field of education in China. Empirical research in this field is increasingly relying on commercial online research groups to collect data. Several data features jointly show that the observed decrease in the consistency of measurement results reflects a systematic response bias, rather than the view that "these constructs are completely consistent in the specific population" proposed in the theoretical hypothesis. Specifically, first of all, in the measurement of the six-point scale, the mean value of all items is centrally distributed in the narrow range of 4.14 to 4.62, and the standard deviation of all items (including those items that are conceptually opposite) is also evenly distributed in the range of 1.45 to 1.73. For example, the average value of the forward statement "AI can help me practice Spanish more efficiently" is 4.40, while the average value of the reverse statement "using AI to complete Spanish homework belongs to cheating behavior" is also as high as 4.45; the two are similar. Secondly, the distribution of respondents' response time is very narrow, with the median of about 200 seconds and the standard deviation of only about 42 seconds, and no respondents can complete all the questionnaires in 100 seconds, which suggests that there may be a

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

fixed, nonreflective pace of answering questions. Third, the missing rate of items in the two independent samples is zero, which is exactly consistent with the online survey design of mandatory answers to all questions, excluding the impact of missing data caused by skipping questions. Fourth, in both samples, the dominant first factor explained more than 49% of the total variance, which was much higher than the empirical threshold set by the traditional Harman single-factor test, further strengthening the possibility of common-method bias.

These features are characteristic of paid online panel data, in which respondents complete surveys quickly and consistently for compensation, and have been documented as risk factors for inflated reliability and inflated factor inter-correlations in Chinese-speaking samples specifically [41, 42]. The implication for the field is significant: as more AI-in-education research relies on commercial Chinese panels for participant recruitment, the risk of measurement artefacts that mimic substantive relationships grows correspondingly. We suggest that future research in this space (a) routinely report Harman's CMB diagnostic and factor inter-correlations alongside conventional fit indices, (b) consider multi-method designs that triangulate self-report data with behavioral or observational indicators, and (c) report recruitment context and respondent compensation structure transparently so that readers can calibrate their interpretation of inferential results.

We note that the present study's descriptive findings (Sections 5.2 and 5.4) are robust to the measurement artefact concerns, because they rely on item-level frequencies and means rather than on between-construct relationships that depend on construct distinctness. The convergent-divergent comparisons reported in Section 5.5 and the province-level patterns reported in Section 5.6 should be interpreted with greater caution and treated as hypothesis-generating rather than confirmatory.

6.4. Pedagogical Implications

Notwithstanding the methodological complications, the descriptive findings support several pedagogical implications worth highlighting. First, given the multi-tool engagement ecosystem documented in Section 5.2, SFL pedagogical interventions should be designed to develop tool-general AI literacy rather than tool-specific skills. Critical evaluation strategies, prompt-design heuristics, and self-monitoring routines should be transferable across the Chinese and international LLM tools that students actually use.

Second, the substantial AI training gap among Spanish teachers identified in Section 5.4 suggests an urgent institutional investment priority. Self-directed teacher learning, while valuable, is insufficient to develop the AI-critical pedagogical competencies described by Alqarni [20] and Wang et al. [36]. Departmental, university, and ministry-level investment in formal teacher GenAI training programmes, tailored to language teaching contexts rather than generic ones, appears warranted.

Third, the student concern about long-term over-reliance, endorsed more strongly than by their teachers, suggests that pedagogical conversations about responsible AI use have an existing student audience receptive to them. The challenge for SFL teachers may be less about persuading students of the risks, a position students already largely hold, than about translating shared concern into concrete strategy use.

6.5. Limitations

Several limitations should be acknowledged. First, the sample was recruited through a commercial online research panel and may not represent the broader population of Chinese university SFL students and teachers in equal measure: in particular, the over-representation of Jiangsu province (approximately one-third of both samples) limits the geographic generalisability of the findings, despite the presence of respondents from 17–18 provinces. Second, the cross-sectional design precludes inference about developmental change in GenAI engagement over time, and the absence of a longitudinal component leaves open the question of how learner engagement patterns evolve with extended GenAI use. Third, as discussed in Section 6.3, the measurement homogeneity observed in the data substantially restricts the inferential conclusions that can be drawn, and the original structural modelling plan could not be

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

executed. Fourth, the absence of a school- or programme-level identifier in the data necessitated the use of province as the proxy clustering variable, which is a coarse aggregate that aggregates considerable institutional heterogeneity. Future research with paired student-teacher data at the programme level would permit a substantially more informative analysis of teacher-student linkage. Fifth, the study relies entirely on self-report and teacher-observer report measures; behavioral and process data (such as the prompt analysis conducted by Huang and Cassany [26]) would substantially enrich understanding of the actual GenAI engagement patterns suggested by the survey data.

7. Conclusion

Through two enlightening core findings, this study has made substantial contributions to the academic discussion on generative AI in the field of language education. First of all, the study focuses on a relatively neglected specific situation - Spanish as a foreign language teaching environment in Chinese universities, and constructs a systematic descriptive analysis framework under this background. The framework deeply analyzes the application mode of generative AI in and out of the classroom, the cognition and attitude of teachers and students, and how these factors affect daily teaching practice. The research findings show that: the student community has spontaneously formed a rich and multi-level artificial intelligence application ecology in practice, and its use strategies and purposes show significant diversity; in contrast, teachers face obvious resource and support gaps in related teaching methods and technical training; it is particularly noteworthy that the students' ability of critical self-reflection in the process of using AI tools is more complex than the simple binary opposition model of "AI-enthusiastic students and AI-cautious teachers", which is often discussed in academia. These solid descriptive research results provide an important empirical basis for further theoretical construction and hypothesis verification in this field.

Second, the study contributes a transparent methodological case study for the rapidly growing body of AI-in-Chinese-education research that depends on commercial online research panels for participant recruitment. The measurement homogeneity observed in our data, theoretically distinct constructs collapsing onto a single response dimension, factor intercorrelations exceeding 0.99, Harman first-factor variance over 49%, is consistent with known features of paid panel data and warrants systematic methodological attention from the field. We hope that the transparent reporting of this measurement outcome will inform both the design and the interpretation of subsequent research and contribute to the development of more robust measurement practices for AI-in-education research at large.

Future research should pursue programme-level paired student-teacher samples through direct institutional recruitment rather than commercial panels; longitudinal designs tracking the development of student GenAI engagement over multiple semesters; and multi-method approaches that combine self-report with behavioral and process data such as prompt analysis or screen recording. The descriptive patterns documented here, particularly the multi-tool engagement ecosystem, the teacher-training infrastructure gap, and the student-teacher divergence on perceived critical engagement, each warrant focused empirical attention with measurement designs that mitigate the bias risks documented in this study.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

© 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

References

[1] X. Huang, D. Zou, G. Cheng, X. Chen, and H. Xie, "Trends, research issues and applications of artificial intelligence in language education," *Educational Technology & Society*, vol. 26, no. 1, pp. 112-131, 2023.

[2] M. Zhu and C. Wang, "A systematic review of research on AI in language education: Current status and future implications," *Language Learning & Technology*, vol. 29, no. 1, pp. 1–29, 2025. https://doi.org/10.64152/10125/73606

[3] B. Wu and X. Yang, "Rethinking language education in the age of generative AI," *Language and Education*, vol. 40, no. 2, pp. 558-561, 2026. https://doi.org/10.1080/09500782.2025.2566796

[4] L. Gutiérrez, "Artificial intelligence in language education: Navigating the potential and challenges of chatbots and NLP," *Research Studies in English Language Teaching and Learning*, vol. 1, no. 3, pp. 180-191, 2023. https://doi.org/10.62583/rseltl.v1i3.44

[5] B. Klimova, M. Pikhart, and J. Kacetl, "ChatGPT and its potential for university foreign language education," *Frontiers in Psychology*, vol. 15, p. 1271, 2024.

[6] H. Çobanoğulları, "Learning and teaching with ChatGPT in foreign language education," *International Journal of Language Education*, vol. 8, no. 2, pp. 45–61, 2024.

[7] E. Creely, "Opportunities and challenges of generative AI for language learning," *Language Learning & Technology*, vol. 28, no. 1, pp. 1–19, 2024.

[8] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025. https://doi.org/10.3390/soc15010006

[9] S. Makeleni, N. Xolisile, and C. Mhaka, "Artificial intelligence in language education: Challenges in the Global South," *Journal of Language Education*, vol. 9, no. 4, pp. 102–118, 2023.

[10] A. M. Mohamed, "Ethical dimensions of AI in language education," *CALL-EJ*, vol. 25, no. 1, pp. 1–22, 2024.

[11] H. Moon, Y. Chung, and A. W. Randolph, "Teaching and learning languages with ChatGPT: Challenges and opportunities in multilingual classrooms in higher education," *Indonesian Journal of English Language Teaching and Applied Linguistics*, vol. 10, no. 1, pp. 207–223, 2025. https://doi.org/10.21093/ijeltal.v10i1.1991

[12] X. Yang and Y. Rui, "ChatGPT for L2 learning: Current status and implications," *System*, vol. 124, p. 103351, 2024. https://doi.org/10.1016/j.system.2024.103351

[13] E. F. Risko and S. J. Gilbert, "Cognitive offloading," *Trends in Cognitive Sciences*, vol. 20, no. 9, pp. 676-688, 2016. https://doi.org/10.1016/j.tics.2016.07.002

[14] R. A. Bjork, "Memory and metamemory considerations in the training of human beings," *Metacognition: Knowing about Knowing*, vol. 185, no. 7.2, pp. 185-205, 1994.

[15] C. Sachdeva and S. J. Gilbert, "Excessive use of reminders: Metacognition and effort-minimisation in cognitive offloading," *Consciousness and Cognition*, vol. 85, p. 103024, 2020. https://doi.org/10.1016/j.concog.2020.103024

[16] J. Iqbal, Z. F. Hashmi, M. Z. Asghar, and M. N. Abid, "Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers," *Scientific Reports*, vol. 15, no. 1, p. 16610, 2025. https://doi.org/10.1038/s41598-025-01676-x

[17] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," *International Journal of Educational Technology in Higher Education*, vol. 20, p. 38, 2023. https://doi.org/10.1186/s41239-023-00408-3

[18] Y. Huang and F. Teng, "Metacognitive and critical thinking skills as prerequisites for effective ChatGPT use in L2 writing," *System*, vol. 131, p. 103744, 2025.

[19] H. Hong, P. Vate-U-Lan, and C. Viriyavejakul, "Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing," *Forum for Linguistic Studies*, vol. 7, no. 7, pp. 325–334, 2025. https://doi.org/10.30564/fls.v7i7.10072

[20] A. Alqarni, "Artificial intelligence-critical pedagogic: Design and psychologic validation of a teacher-specific scale for enhancing critical thinking in classrooms," *Journal of Computer Assisted Learning*, vol. 41, no. 3, p. e70039, 2025. https://doi.org/10.1111/jcal.70039

[21] Y. Gu and W. N. E. H. Sharil, "Language learning strategies, motivation and EFL proficiency: A study of Chinese tertiary-level non-English majors," *Eurasian Journal of Applied Linguistics*, vol. 10, no. 2, pp. 254-261, 2024.

[22] A. Habók, Y. Kong, J. Ragchaa, and A. Magyar, "Cross-cultural differences in foreign language learning strategy preferences among Hungarian, Chinese and Mongolian University students," *Heliyon*, vol. 7, no. 3, p. e06505, 2021. https://doi.org/10.1016/j.heliyon.2021.e06505

[23] W. Zhangli, W. Xuewen, L. Ismail, N. Khaiessa Ahmad, and O. H. Mahfoodh, "Challenges and strategies in English-medium learning among Chinese international students in a Malaysian university," *SAGE Open*, vol. 15, no. 2, p. 21582440251341084, 2025. https://doi.org/10.1177/21582440251341084

[24] S. Zhou and Y. Zhang, "Developmental trajectories of and reciprocal relationships between Chinese university students' foreign language reading self-efficacy and reading strategy use," *Frontiers in Psychology*, vol. 16, p. 1512098, 2025. https://doi.org/10.3389/fpsyg.2025.1512098

[25] G. P. Georgiou, "ChatGPT produces more "lazy" thinkers: Evidence of cognitive engagement decline," *arXiv:2507.00181 [Preprint]. arXiv*, 2025. https://doi.org/10.48550/arXiv.2507.00181

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 7: 47-66, 2026
DOI: 10.55214/2576-8484.v10i7.13210
© 2026 by the authors; licensee Learning Gate

[26] S. Huang and D. Cassany, "Spanish language learning in the AI era: AI as a scaffolding tool," *Journal of China Computer-Assisted Language Learning*, vol. 5, no. 1, p. Article e26, 2025. https://doi.org/10.1515/jccall-2024-0026

[27] M. Vendrell and S.-K. Johnston, "Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education," *Computers and Education: Artificial Intelligence*, vol. 10, p. 100572, 2026. https://doi.org/10.1016/j.caeai.2026.100572

[28] R. Toncelli and I. Kostka, "A love-hate relationship: Exploring faculty attitudes towards GenAI and its integration into teaching," *International Journal of TESOL Studies*, vol. 6, no. 3, pp. 77–94, 2024. https://doi.org/10.58304/ijts.20240306

[29] S. Zaimoğlu and A. Dağtaş, "Teacher cognition and practices in using generative AI tools to support student engagement in EFL higher-education contexts," *Behavioral Sciences*, vol. 15, no. 9, p. 1202, 2025. https://doi.org/10.3390/bs15091202

[30] M. Pan, C. Lai, and K. Guo, "Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading," *The Internet and Higher Education*, vol. 65, p. 100991, 2025. https://doi.org/10.1016/j.iheduc.2024.100991

[31] X. Wang, Y. Gao, Q. Wang, and P. Zhang, "Fostering engagement in AI-mediated Chinese EFL classrooms: The role of classroom climate, AI literacy, and resilience," *European Journal of Education*, vol. 60, no. 1, p. e12874, 2025. https://doi.org/10.1111/ejed.12874

[32] D. T. K. Ng, C. W. Tan, and J. K. L. Leung, "Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study," *British Journal of Educational Technology*, vol. 55, no. 4, pp. 1328-1353, 2024. https://doi.org/10.1111/bjet.13454

[33] H. Yu, Y. Guo, H. Yang, W. Zhang, and Y. Dong, "Can ChatGPT revolutionize language learning? Unveiling the power of AI in multilingual education through user insights and pedagogical impact," *European Journal of Education*, vol. 60, no. 1, p. e12749, 2025. https://doi.org/10.1111/ejed.12749

[34] Y. Li and Q. Zhang, "Exploring the impact of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment," *Heliyon*, vol. 10, no. 21, p. e39845 2024.

[35] B. Cabellos, C. de Aldama, and J.-I. Pozo, "University teachers' beliefs about the use of generative artificial intelligence for teaching and learning," *Frontiers in Psychology*, vol. 15, p. 1468900, 2024. https://doi.org/10.3389/fpsyg.2024.1468900

[36] Y. Wang, A. Derakhshan, and F. Ghiasvand, "EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study," *System*, vol. 133, p. 103791, 2025. https://doi.org/10.1016/j.system.2025.103791

[37] Z. Xingzun and W. Hongtao, "Metabolite changes during developmental transitions in Adonis amurensis Regel et Radde flowers: Insights from HPLC-MS analysis," *PLOS One*, vol. 20, no. 1, p. e0313337, 2025. https://doi.org/10.1371/journal.pone.0313337

[38] S. De Vries, W. J. van de Grift, and E. P. Jansen, "How teachers' beliefs about learning and teaching relate to their continuing professional development," *Teachers and Teaching*, vol. 20, no. 3, pp. 338-357, 2014. https://doi.org/10.1080/13540602.2013.848521

[39] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, pp. 115-135, 2015. https://doi.org/10.1007/s11747-014-0403-8

[40] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: A critical review of the literature and recommended remedies," *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879-903, 2003. https://psycnet.apa.org/doi/10.1037/0021-9010.88.5.879

[41] A.-W. Harzing, "Response styles in cross-national survey research: A 26-country study," *International Journal of Cross Cultural Management*, vol. 6, no. 2, pp. 243-266, 2006. https://doi.org/10.1177/1470595806066332

[42] J. P. Robinson-Cimpian, "Inaccurate estimation of disparities due to mischievous responders: Several suggestions to assess conclusions," *Educational Researcher*, vol. 43, no. 4, pp. 171-185, 2014. https://doi.org/10.3102/0013189X14534297

© 2026 by the authors; licensee Learning Gate