Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

High-Performance Configurations in Game-Based English Learning: How AI Support and Task Complexity Matter

Dongling Ma
Wuhan Institute of Technology and Business
Wenjing Zhou
2006094@wtu.edu.cn
Wuhan Textile University
Weizheng Jiang
Wuhan Institute of Technology and Business

Article

Keywords: game-based learning (GBL), English as a foreign language (EFL), task complexity, learning behavior configurations, fuzzy-set qualitative comparative analysis (fsQCA)

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11382671/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

High-Performance Configurations in Game-Based English Learning: How AI Support and Task Complexity Matter

Dongling Ma[1], Wenjing Zhou[2*], Weizheng Jiang[1]

[1]School of Management, Wuhan Institute of Technology and Business, Wuhan, China

[2] School of Foreign Studies, Wuhan Textile University, Wuhan, China

Email: 2006094@wtu.edu.cn

Abstract

Artificial intelligence is not only empowering education but also challenging traditional educational designs and learning models. Against this backdrop, game-based learning needs to pay greater attention to how different learning behaviors adapt to increasingly complex tasks to facilitate high task performance. This study focuses on EFL learning in higher education and develops a configurational framework based on internally driven conditions and external support conditions in GBL-based EFL learning. To compare changes in learning-behavior configurations across different task stages, 33 students completed three tasks with low, medium, and high levels of complexity. Fuzzy-set qualitative comparative analysis (fsQCA) was then employed to identify behavioral configurations associated with high task performance. The results show that, in an AI-scaffolded learning environment, the progressive change in task complexity was accompanied by changes in the number of learning-behavior configurations. As task complexity increased, the configurational roles of internally driven conditions and AI scaffolding conditions became increasingly differentiated within high-performance configurations. These findings suggest that GBL-based EFL performance does not depend on any single learning behavior but rather on the dynamic configuration of learner behaviors and external support conditions in response to task demands.

Key words： game-based learning (GBL); English as a foreign language (EFL); task complexity; learning behavior configurations; fuzzy-set qualitative comparative analysis (fsQCA)

1. Introduction

Generative artificial intelligence is reshaping higher education, and under this trend, tasks in English as foreign language (EFL) courses are gradually shifting toward more contextualized, interactive, and practice-oriented learning environments[1]. Meanwhile, with the emergence of large language models, courses related to foreign-language professional skills have increasingly emphasized the in-depth application of these skills within specific domains and the development of cross-cultural communication competence. In China, English, as a specialized professional skill, is increasingly being integrated and combined with other disciplines. An important objective of this integration is to strengthen the instrumental and cross-cultural functions of English and thereby enhance its application to problem-solving across different industries[2]. Against this backdrop, game-based learning (GBL) provides a relatively suitable practical environment for vocational EFL instruction. A key feature of GBL is its ability to embed language knowledge within task contexts characterized by goals, rules, feedback, and role-based interaction, allowing learners to learn through action, experimentation, and feedback. Previous studies have shown that the themes, complexity, and interaction design of game-based

tasks can significantly influence the performance of EFL learners[3].

However, existing literature and empirical studies have revealed considerable debate regarding the role of AI in GBL-EFL learning environments. At the same time, the high transparency and functionality of AI have raised a question that remains insufficiently addressed[4]. In GBL-EFL learning environments, how should self-directed learning behaviors and external support be configured to maintain high learning performance across tasks of different levels of complexity? In traditional GBL environments, learners primarily advance their learning through personal engagement, repeated practice, and feedback-based correction, while obtaining social support through peer interaction. However, in the current learning environment, learners inevitably need to confront the double-edged effects of generative AI[5]. Related studies have begun to examine the relationship between AI and self-regulated learning, as well as how AI can be appropriately embedded as scaffolding within everyday learning activities[6].

A major limitation of existing research is that it tends to adopt a static perspective to explain the effects of individual factors on learning performance. GBL places greater emphasis on process-oriented learning, in which dynamic processes such as feedback, interaction, and action constitute key characteristics of learning activities. Individual behavioral characteristics and static outcome measures cannot fully reveal the underlying mechanisms through which learning performance emerges. More importantly, these dynamic processes are accompanied by participants' continuous adjustments to their learning strategies and access to learning resources. Therefore, simply determining which strategy is more effective in GBL may not provide a sufficiently in-depth understanding of the complex behavioral processes involved in EFL learning in the AI era.

To address this issue, this study employs fsQCA to examine behavioral changes throughout the GBL process. Drawing on self-regulated learning (SRL), experiential learning, and scaffolding theory, we identify both internal and external conditions within the GBL process, including Engagement, Practice Feedback, Peer Interaction, and AI Support. fsQCA is then used to identify different configurations through which these conditions jointly produce high learning performance across multiple tasks. The study involves 33 university students and incorporates three progressive vocational-context tasks into a tourism English course. To capture the processual and progressive nature of learning, the three tasks were designed with gradually increasing levels of complexity. In the data analysis, we separately identify the configurations associated with high-efficiency learning performance for each task and compare how these configurations are reconfigured as task complexity increases.

By integrating fsQCA with progressive task contexts, this study examines high learning performance in GBL as a result of multiple interacting factors. Our focus is on how different learning behaviors and external supports form task-dependent configurations in AI-supported

GBL-EFL environments and how these configurations change with increasing task complexity. This perspective contributes to a deeper understanding of the dynamic relationships among learner-initiated behaviors, peer interaction, and AI support in vocational EFL learning in the AI era. It also provides a basis for designing GBL instruction and configuring AI scaffolding across tasks with different levels of complexity.

2. Theoretical Background and Research Framework

2.1 Learner-driven learning processes

Game-based learning does not emerge independently of existing educational theories; rather, it represents a continuation of the development of established learning theories, while placing greater emphasis on learners' agency. These theories provide GBL with a multidimensional and dynamic perspective rooted in learners' active engagement in the learning process. Task orientation is an important characteristic of GBL, and learners' attitudes, actions, and decisions toward a task may all have important implications for learning performance. Previous research has also established strong connections between GBL and both experiential learning theory and self-regulated learning theory.

Kolb's experiential learning theory proposes that the cyclical process of engagement, reflection, conceptualization, and experimentation provides learners with a cognitive framework and a model for action[7]. The theory emphasizes learners' subjective experiences and highlights the continuous revision and reconstruction of knowledge throughout the dynamic learning process[8].

The rapid development of artificial intelligence in recent years makes self-regulated learning (SRL) an important factor to consider in the present study. SRL emphasizes learners' active implementation, monitoring, and adjustment of their own learning behaviors[9]. A review of recent studies reveals a relatively consistent view[6,e.g., 10–12]: in the AI era, learning outcomes do not simply depend on whether learners use AI, but rather on how they integrate their own learning agency with external AI support.

Building on these theoretical perspectives, this study focuses on three interrelated process variables that may exhibit nonlinear relationships: Engagement, Practice, and Feedback. Engagement represents the initial stage of the experiential learning cycle and provides the foundation for learners to invest attention, time, and behavioral effort in active practice. Practice is closely associated with reflection and observation, through which learners can test existing knowledge and learning strategies and adjust them in response to task-related feedback. Active Feedback, in turn, constitutes an important stage in the development of abstract conceptualization, as it involves learners' active attention to, understanding of, and uptake of

feedback, which subsequently informs adjustments to their information-processing and learning strategies.

Within GBL tasks, the relative contribution of behavioral characteristics such as task engagement, action-based practice, and Feedback may not be equivalent. Particularly as task complexity increases, learners may need to adjust their overall approach to learning as well as the configuration of different learning behaviors.

2.2 Learning scaffolding

Scaffolding is commonly used to describe the processes of providing support within the zone of proximal development, constructing temporary support structures, and gradually withdrawing them as learners become more capable of independent performance. Previous research suggests that scaffolding can take various forms, including instructional design, visualization, prompting techniques, and peer assistance. In GBL-based EFL higher education contexts, the careful design of scaffolding can be facilitated through peer support [13]. EFL teaching and learning in higher education does not follow a fixed structure or produce predetermined outcomes. Scaffolding is therefore needed to foster collective critical thinking and to reduce the constraints imposed by established learning patterns on learners [14]. Meanwhile, Hammond's research on language education has placed particular emphasis on how instructional design can enhance students' affective engagement [15]. Thus, the essence of scaffolding does not lie simply in "providing help"; rather, its effectiveness may vary with changes in the learning context, task requirements, and learner states.

In GBL environments, peer interaction constitutes an important form of scaffolding. The role of peers does not necessarily involve directly providing the final answer; instead, it is manifested more prominently in task resolution and cognitive construction throughout the learning process. From the perspective of Bloom's taxonomy, peer scaffolding can play an important role at higher-order cognitive levels, such as commenting, analyzing, and creating [16].

Unlike conventional digital tools, AI not only serves a tool-like function but is increasingly becoming an intelligent and dynamic form of scaffolding across different learning contexts. AI is reshaping the allocation of human cognitive labor and can take on a substantial role in processing complex information and performing basic information-organizing tasks, thereby allowing humans to devote their limited cognitive resources to higher-order activities such as critical thinking and decision-making [4,17]. However, within GBL, the role of AI as process-oriented scaffolding still requires further examination and development. For example, Ma and Chen's EFL study suggests that when AI is effectively integrated with human knowledge, it can help address limitations related to cultural nuances, overcorrection, and trust. Previous studies have also suggested [17,18] that, as AI becomes increasingly integrated into and gradually

incorporated into traditional education, AI alone may not be sufficient to provide independent learning guidance. Therefore, the learning value of AI support still depends on how learners select, interpret, and define the learning resources available to them.

In this study, scaffolding theory is further conceptualized as a form of **external learning support**, without presuming that such support will necessarily lead to positive learning outcomes. Within the context of the present study, this external support is operationalized as **Peer Interaction** and **AI Support**.

2.3 Theoretical framework

Building on experiential learning theory, self-regulated learning theory, and scaffolding theory, this study further develops a task-contingent configurational framework that incorporates five conditions of the learning process into a unified analytical framework: Engagement, Practice, Feedback, Peer Interaction, and AI Support (see **Figure 1**).

Engagement, Practice, and Feedback represent different manifestations of learners' active involvement in driving the learning process and are conceptualized as **internally driven conditions**. In contrast, Peer Interaction and AI Support represent **external support conditions**, providing information, perspectives, strategies, or problem-solving cues through social interaction and intelligent technologies, respectively. Together, these two groups of conditions characterize the GBL process from complementary perspectives and are theoretically complementary to each other.

More importantly, the five conditions may not have a unique or fixed matching relationship across tasks of different difficulty levels. As task complexity increases, learners need to continuously adjust and improve their cognitive strategies and approaches to information integration, which in turn requires corresponding changes in their behavioral configurations. Therefore, based on theoretical synthesis and deduction, this study proposes that, in changing EFL learning contexts, different conditions can combine in different ways to form **effective learning configurations**.

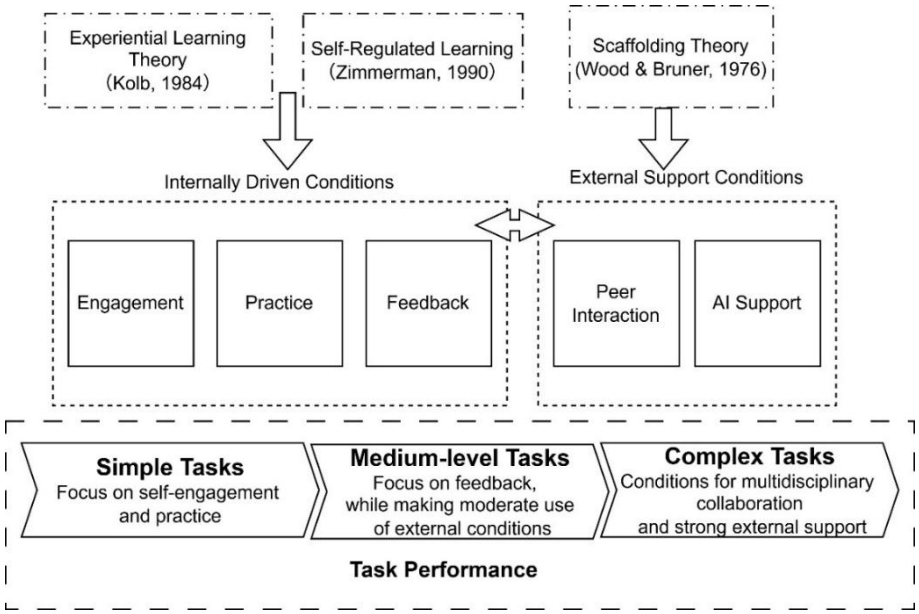


Figure 1 The theoretical framework of the study

2.4 Research questions

This study does not presuppose the effect of any single condition on task performance. Instead, it adopts a configurational perspective to examine the learning configurations associated with high task performance across different levels of task complexity. Accordingly, the following research questions are proposed:

RQ1: In game-based EFL learning across different levels of task complexity, which configurations of internally driven behaviors and external support are associated with high task performance?

RQ2: How do these high-performance configurations change as task complexity increases?

3 Methods

3.1 Research design and participants

This study was conducted within the *Tourism English* course in the Tourism Management program at Wuhan Institute of Technology and Business. The course is an EFL course in higher education and a compulsory course for students majoring in Tourism Management. It is characterized by a strong emphasis on practical application, vocational competence, and comprehensive skill development. The primary purpose of the course is to establish common vocational task contexts and develop students' English communication, cross-cultural communication, and foreign-related service competencies in tourism settings. Unlike general EFL courses, this course emphasizes the construction and enrichment of learning contexts

through practical scenario-based tasks. Through laboratory-based learning environments, VR technology, and simulated professional roles, learners develop competencies in information processing, communicative expression, and problem solving.

The participants were 33 students enrolled in the 2025 cohort of the undergraduate top-up program who were taking the course. They were aged between 21 and 24 years, with females accounting for approximately 73% of the sample. All participants had a certain foundation in English learning and vocational practice. Before entering the undergraduate program, they had received approximately three years of English education at the junior-college level and had participated in relevant industry practice. At the undergraduate level, EFL learning as part of professional skills development places greater emphasis on professional application and vocational competence. The preparation phase of the study took place from October 2025 to January 2026 and included a pilot study involving similar scenario-based tasks. The main study was conducted within the course from February to June 2026. Participants completed three GBL - EFL learning tasks. All three tasks were situated in tourism-related vocational contexts and increased progressively in terms of task requirements and cognitive processing difficulty. The students' behavioral processes across the three tasks were quantitatively recorded, resulting in 99 task-level observations, which were subsequently analyzed separately using fsQCA.

3.2 Game-based learning design

This study designed three progressive GBL tasks: **Food and Beverage, Shopping for Souvenirs,** and **Sightseeing**, corresponding to low-, medium-, and high-difficulty tasks, respectively (see **Table 1**). Previous research has shown that instructional tasks with different levels of complexity can stimulate group dynamics and learning immersion, not only significantly improving learning outcomes[19] but also influencing learners' behavioral performance in tasks through cognitive enhancement [20]. Based on these theoretical and practical considerations, this study designed three vocational scenario-based tasks with different levels of difficulty, requiring participants to complete the tasks through information acquisition, communicative interaction, and contextual decision-making.

Task 1 (Food and Beverage) was designed as a low-difficulty task and primarily involved describing and recommending dishes. The information involved was relatively limited, and the decision-making requirements were low. The vocational interaction in this task was relatively straightforward and mainly involved remembering and understanding. Participants could complete the scenario-based task by drawing on existing vocabulary and language expressions.

Task 2 (Shopping for Souvenirs) required participants to compare, select, and bargain over products, thereby integrating multiple types of information to make choices and decisions. This task was designed as a medium-difficulty task and required participants to move beyond the direct constraints of course content and flexibly apply their acquired knowledge to solve

problems.

Task 3 (Sightseeing) was designed as a high-difficulty task. It not only required participants to take on the role of a tour guide but also incorporated multiple constraints into a complex travel-planning process, including time, routes, and budget. In addition, explaining tourist attractions went beyond language learning itself and involved cross-cultural communication. Therefore, this task involved the highest level of integrative and cognitive demands.

Table 1 Three Different-Difficulty Scenario Tasks

Task	Context	Content	Complexity level
T1: Food and Beverage	Ordering food in the restaurant and recommending dishes	Role-play: Introduction and Recommendation of Dishes	low
T2: Shopping for Souvenirs	Choosing and purchasing souvenirs for tourists	Role-play: Comparison and introduction of products, bargaining	Moderate
T3: Sightseeing	Plan and guide tourist destinations for visitors	Arrangement design: Time limit, route planning, budget constraints; Role-playing: Tourist attraction explanation	Difficulties

3.3 Behavioral measures and data coding

Because questionnaire-based measurement relies heavily on participants' subjective experiences and attitudes, it may compromise the objectivity of data analysis. Therefore, this study further quantified behavioral trace data generated during the learning process [21] (see **Table 2**). Previous studies have demonstrated the reliability and validity of this approach. For example, Alhazbi et al. [22] extracted students' self-regulated learning behaviors from online tracking data, while Du et al. [21] and Ye and Pennisi [23] adopted similar approaches. The procedures and findings of these studies further highlighted the need to establish objective and observable learning samples. In the present study, the 33 participants served as case carriers of behavioral trace data, and their behaviors during EFL-GBL learning were operationalized as combinations of conditions in the fsQCA. This approach enables a more direct characterization of configurational differences in learning behaviors across different task contexts.

In fsQCA, cases constitute the basic units of configurational analysis. Therefore, unlike conventional regression or structural equation modeling, fsQCA does not primarily aim to estimate parameters using large samples; instead, it focuses on the combinations of conditions across cases and their set-theoretic relationships with the outcome. Previous fsQCA research

has shown that configurational analysis can still be conducted with a relatively limited number of cases when frequency thresholds, consistency thresholds, and other truth-table parameters are appropriately specified [24]. In the present study, fsQCA was conducted using behavioral data from 33 participants, while the 99 task-level observations generated across the three tasks were used to characterize learning behaviors at different task stages. This design preserves differences in individual behavioral configurations while avoiding treating repeated observations from the same participant across different tasks as statistically independent cases.

Engagement was used to represent the extent to which students participated in the tasks, as measured by the number of classroom and online participation instances. This indicator captures learners' actual engagement behaviors rather than their subjective perceptions and is therefore suitable for objectively assessing participants' task engagement [25]. **Practice** was used to reflect the extent of learners' continued practice and repeated attempts, operationalized as the ratio of the number of actual attempts completed to the total number of attempts allowed or required. **Feedback** (or Feedback Utilization) was measured through the relationship between the number of corrections and the number of correctable errors, emphasizing the actual use of feedback rather than its mere receipt [26]. It should be noted that most of the raw data for **Engagement, Practice, and Feedback** were obtained from the Xuexitong mobile application.

Peer Interaction was quantified by constructing a consultation network among students [27] and using each individual's centrality within the overall network to represent the extent of peer interaction. Previous studies have shown that individual centrality is an important behavioral indicator of innovation-driven activity and is associated with learning performance [16,28]. **AI Support** was quantified by coding AI dialogue logs to reflect the extent to which learners actually used AI support. This measurement approach was informed by recent research in higher education, in which interaction logs generated by online systems were used to directly analyze students' actual behavioral engagement rather than relying on self-reported measures [23]. In the present study, the scope of AI use was further defined: large language models (LLMs) were limited to the Doubao application, and the functional boundaries were restricted to the existing functions and capabilities provided by the application.

Task Performance, as the outcome variable, represents the assessment of learning outcomes following task completion. It consisted of a knowledge test (**post-test**) and a **scenario-based transfer report**. Previous studies support this approach to measurement [29,30]. Specifically, the knowledge test assessed learners' mastery of the target knowledge, whereas the scenario-based transfer report assessed their ability to transfer what they had learned to a new context. The corresponding measurement procedures are provided in **Appendix A**.

Table 2: The Connotations and Quantification of Variables

Variables	Behavioral indicators	Data source	Operationalization approach
Engagement	number of proactive attempts	game logs, classroom observations, online records	proactive attempts count
Practice	completion rate of exercises		completed / total
Feedback	feedback utilization		correctable behavior/correctable errors
Peer Interaction	The centrality of peer interaction	nomination method	advice network centrality
AI Support	AI usage patterns	AI chat log	AI interaction count
Task Performance	task performance	post-test、transfer report	task performance score

3.4 Calibration, fsQCA procedure, and cross-task comparison

This study employed fuzzy-set qualitative comparative analysis (fsQCA) to identify multiple configurations of learning conditions associated with high-efficiency learning performance. All variables in this study were derived from observations of participants' actual behaviors rather than questionnaire-based measures. Therefore, the data were calibrated using the K-means clustering method, a machine learning approach, to determine the thresholds for fuzzy-set calibration. This threshold-setting approach has also been applied in our previous studies [17,31]. Specifically, we first calculated the within-cluster sum of squares (WCSS) across different clustering solutions and identified the clustering solution that minimized the WCSS, as shown in Equation (1):

$$\min_{C} \sum_{k=1}^{3} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2 \quad (1)$$

Here, $X = \{x_1, x_2 \cdots, x_n\}$ represents the performance dataset of the participants, C_k denotes the k-th cluster, and μ_k represents the center of the k-th cluster, that is, the mean of all observations within that cluster. The generalized logistic function was then used to transform the original variable (x) into a fuzzy-set membership score, as shown in Equation (2) [32]. In this function, α controls the slope of the logistic curve.

$$f(x) = \frac{1}{1 + \exp(-\alpha(x - x_{0.5}))} \quad (2)$$

For different types of indicators, including behavioral frequencies, proportions, and network centrality, the corresponding calibration thresholds were determined based on the characteristics of the data distributions and the K-means calibration principle. Direct calibration was then used to transform the raw data into fuzzy-set membership scores ranging from 0 to 1. Task performance, as the outcome variable, was calibrated based on its actual score distribution to identify cases exhibiting high-efficiency learning performance. The calibration anchors for all variables included full membership, the crossover point, and full non-membership, as presented in **Table 3**. Table 3 further presents the datasets for the three tasks with different levels of difficulty and their corresponding calibration thresholds.

Table 3 Calibration of Variables

Variables	full non-membership	crossover point	full membership
		Task 1	
Engagement	3.25	5.64	14.5
Practice	0.33	0.62	0.88
Feedback	0.11	0.57	0.85
Peer Interaction	8.6	13.18	20.75
AI Support	10.55	16.64	21.88
Task Performance	76.66	80.17	84.97
		Task 2	
Engagement	3.5	5.46	10.5
Practice	0.37	0.63	0.92
Feedback	0.47	0.66	0.88
Peer Interaction	9.09	15.75	23.5
AI Support	10.54	16.1	21.4
Task Performance	65.64	74.98	81.95
		Task 3	
Engagement	2.69	4.47	8.33
Practice	0.34	0.63	0.89
Feedback	0.21	0.59	0.87
Peer Interaction	9.67	16.8	24.18
AI Support	13.5	19.5	33.8
Task Performance	62.56	79.32	84.1

This study conducted fsQCA for simple, moderate, and difficult tasks, respectively. Unlike most studies, this research did not combine the 99 observation samples from the three tasks into a single analysis. To identify the stabilization, replacement, and restructuring of efficient learning paths as task complexity increases, this study analyzed and compared the three sets of fsQCA results, including core conditions, peripheral conditions, and learning paths.

3.5 Theoretical basis of configurational conditions

The conditions examined in this study were theoretically specified based on the framework

described above. The internally driven conditions were primarily derived from experiential learning theory and self-regulated learning theory, whereas the external support conditions were mainly derived from scaffolding theory. Existing research has provided some empirical support for this approach. For example, Dong et al. [33] examined the relevant theoretical relationships using SEM and fsQCA, incorporating technological, interactional, and competence-related factors into a unified theoretical framework and using fsQCA to identify multiple configurations associated with high levels of engagement. A similar approach was adopted in a more recent study by Zhao and Zhang [34]. Accordingly, this study does not presuppose that any single behavior has a stable net effect across different tasks. Instead, building on established theoretical perspectives, the above behavioral conditions are incorporated into a unified configurational framework to examine how they combine in different ways to produce high levels of task performance.

It should also be noted that cases in this study were defined according to learners' actual behaviors during the tasks rather than demographic characteristics such as gender or age. Each participant constituted an individual case. In fsQCA, the focus is on cases' set memberships in behavioral conditions and the combinations of these conditions, rather than on grouping cases according to demographic attributes [35,36]. Accordingly, demographic characteristics were not specified as conditions in the configurational analysis.

3.6 Ethical considerations and funding

This study received ethical approval from the Research Department of Wuhan Institute of Technology and Business. The study was supported by the Hubei Province Higher Education Institutions' Outstanding Middle-Aged and Young Scientific and Technological Innovation Teams Project (Grant No. T2022049) and the Hubei Province Higher Education Institutions' Clusters of Distinctive and Strength-Based Disciplines - "Digital Business and Management" Project (Grant No. ACDG202501). The study was conducted in accordance with relevant research ethics standards. Before the study commenced, participants were informed of the research purpose, data collection procedures, and their rights as participants, and they participated in the study after providing informed consent. Classroom behavioral records, interaction data, and AI dialogue data collected in the study were used solely for research purposes and were anonymized during data processing to protect participants' privacy. The findings are reported only in aggregated form, and no information that could identify individual participants is disclosed.

4. Result

4.1 Descriptive Statistics

Descriptive statistics were calculated for the three tasks, and the results are presented in **Table 4**. Overall, as task complexity increased, the means of the external support conditions showed an increasing trend, whereas the internally driven conditions remained relatively stable. At the same time, the dispersion of learning outcomes increased.

Regarding the internally driven conditions, the mean values of Engagement remained relatively stable across the three tasks, at 4.73, 4.70, and 4.61, respectively, indicating that the number of learners' active attempts did not increase substantially as task complexity increased. The mean values of Practice were 0.583, 0.612, and 0.602 for Tasks 1, 2, and 3, respectively, remaining relatively stable overall. In contrast, Feedback showed greater variation, increasing from 0.607 in Task 1 to 0.675 in Task 2 and then decreasing to 0.614 in Task 3.

Regarding the external support conditions, Peer Interaction showed a gradual increasing trend, with its mean increasing from 14.55 in Task 1 to 15.88 in Task 2 and further to 16.67 in Task 3. The mean values of AI Support were relatively similar in Tasks 1 and 2, at 15.88 and 15.52, respectively, but increased substantially to 19.85 in Task 3.

Regarding the outcome variable, the mean value of Task Performance decreased slightly from T1 to T2 and remained broadly stable in T3, whereas its standard deviation (SD) increased. This indicates that individual differences in learning outcomes became more pronounced in the high-difficulty task.

Table 4. Descriptive statistics across the three tasks

Variable	Task 1 Mean (SD)	Task 2 Mean (SD)	Task 3 Mean (SD)
Engagement	4.73 (2.92)	4.70 (1.90)	4.61 (3.33)
Practice	0.583 (0.211)	0.612 (0.230)	0.602 (0.240)
Feedback	0.607 (0.219)	0.675 (0.178)	0.614 (0.211)
Peer Interaction	14.55 (5.42)	15.88 (6.15)	16.67 (6.50)
AI Support	15.88 (4.64)	15.52 (4.81)	19.85 (7.06)
Task Performance	80.63 (3.37)	79.90 (4.82)	79.97 (6.42)

4.2 Necessity Analysis

The primary focus of this study is on the configurations associated with high-efficiency learning performance (fsTask Performance). **Table 5** presents the necessity analysis results for fsTask Performance, while the corresponding results for ~fsTask Performance (low-efficiency learning performance) are provided in the **Appendix Table A2**.

Across all three tasks, none of the individual conditions reached the commonly adopted

consistency threshold of 0.90 for a necessary condition. In Task 1, ~fsEngagement showed the highest consistency with fsTask Performance at 0.898, but this value remained below 0.90. In Task 2, ~fsPractice showed a consistency of 0.943 with ~fsTask Performance, whereas its consistency with fsTask Performance was only 0.542. In Task 3, the consistencies of ~fsEngagement and ~fsPractice with ~fsTask Performance were 0.948 and 0.937, respectively, while their corresponding consistencies with fsTask Performance were only 0.711 and 0.888.

Table 5 Analysis of the Necessity of fsTask Performance (High-Efficiency Learning Performance)

Condition	Task 1	Task 2	Task 3
fsEngagement	0.337	0.307	0.456
~fsEngagement	0.898	0.785	0.711
fsPractice	0.505	0.539	0.192
~fsPractice	0.766	0.542	0.888
fsFeedback	0.723	0.529	0.658
~fsFeedback	0.552	0.538	0.513
fsPeer Interaction	0.63	0.511	0.507
~fsPeer Interaction	0.616	0.572	0.6
fsAI Support	0.56	0.488	0.503
~fsAI Support	0.694	0.597	0.678

Overall, no single condition or its absence was found to consistently constitute a necessary condition for high-efficiency learning performance across the three tasks. Therefore, the subsequent configurational analysis does not treat any single condition as a necessary prerequisite for fsTask Performance. Instead, it examines how multiple conditions combine in different ways to form configurations associated with fsTask Performance.

4.3 Cross-task trajectories

Building on the necessity analysis, fsQCA was further employed to identify the configurations associated with fsTask Performance across different levels of task complexity. The configurational analysis identified three configurations for the low-difficulty task, seven configurations for the medium-difficulty task, and four configurations for the high-difficulty task. As task complexity changed, the combinations of learner behaviors and external scaffolding conditions were substantially reconfigured, indicating that high task performance could be achieved through multiple configurations rather than through a single configuration. In terms of the configurational measures, raw coverage ranged from 0.149 to 0.448, unique coverage ranged from 0.023 to 0.201, and consistency ranged from 0.802 to 0.996.

4.3.1 Low-Difficulty Task

For the low-difficulty task, three configurations were identified (**Table 6**). **Configuration 1 (C1)** was characterized by the simultaneous presence of Engagement, Practice, Feedback, and

Peer Interaction, while AI Support was absent from the configuration. The raw coverage was 0.251, the unique coverage was 0.150, and the consistency was 0.802. This configuration indicates that, under relatively low task complexity, a combination of high levels of active Engagement, Practice, and Feedback with peer interaction can be associated with high Task Performance.

C2 exhibited a substantially different behavioral combination, characterized by the absence of Engagement and Practice, the presence of Feedback, and the absence of both Peer Interaction and AI Support. Its raw coverage was 0.359, unique coverage was 0.201, and consistency reached 0.843. Compared with the other configurations, C2 had relatively high raw and unique coverage in the low-difficulty task.

C3 was characterized by the absence of Engagement, Practice, and AI Support, while Feedback and Peer Interaction were present. Its raw coverage was 0.248, unique coverage was 0.102, and consistency was 0.812.

Table 6 Configuration of Simple Tasks (fsTask Performance)

Conditional Variables	**C1**	**C2**	**C3**
Engagement	●	⊗	⊗
Practice	●	⊗	⊗
Feedback	●	●	⊗
Peer Interaction	●	⊗	●
AI Support		⊗	⊗
Raw coverage	0.251	0.359	0.248
Unique coverage	0.15	0.201	0.102
Consistency	0.802	0.843	0.812

Note: ● = presence of a core condition; ⊗ = absence of a core condition; ○ = presence of a peripheral condition; ∅ = absence of a peripheral condition; blank = the condition is not part of the configuration. The results of the robustness test are provided in **Appendix B**.

Overall, the low-difficulty task exhibited a relatively clear pattern of multiple configurations associated with high task performance. On the one hand, high task performance could emerge from combinations of proactive learning behaviors and effective social interaction. On the other hand, when AI Support did not play a prominent role in the low-difficulty task, Engagement, Practice, Feedback, and Peer Interaction appeared in different combinations across the high-performance configurations, suggesting a certain degree of potential substitutability among these conditions.

4.3.2 Medium-Difficulty Task

Seven configurations were identified for the medium-difficulty task (**Table 7**), making it the most configurationally complex stage among the three tasks. The consistency values of all

configurations were relatively high, ranging from 0.918 to 0.996. However, the unique coverage values were relatively limited, ranging from 0.023 to 0.074. The combination of high consistency and relatively low unique coverage suggests that high task performance was not explained by a single dominant configuration but was jointly covered by multiple, relatively dispersed combinations of conditions.

In the medium-difficulty task, **Configuration 1 (C1)** and **Configuration 3 (C3)** had relatively high coverage. In C1, Practice was a core present condition, while Engagement was a peripheral absent condition. Its raw coverage was 0.444 and its consistency was 0.980, making it the configuration with the highest coverage for this task. C3 was characterized by the peripheral absence of Engagement and the core presence of Feedback and Peer Interaction. Its raw coverage was 0.323 and its consistency was 0.952. The remaining configurations had relatively lower coverage, further demonstrating the diverse combinational relationships among the conditions.

Table 7 Configuration for Medium-Difficulty Tasks (fsTask Performance)

Conditional Variables	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
Engagement	⊘		⊘		○	⊘	
Practice	●	⊘		●	⊘		●
Feedback		○	●			⊗	⊘
Peer Interaction			●	⊘	⊘	⊗	○
AI Support		⊗		○	⊗	○	⊗
Raw coverage	0.444	0.278	0.323	0.256	0.149	0.197	0.192
Unique coverage	0.074	0.048	0.029	0.043	0.038	0.047	0.023
Consistency	0.98	0.918	0.952	0.996	0.973	0.973	0.994

Note: The results of the robustness test are provided in **Appendix C**.

Therefore, as a transitional task, the medium-difficulty task did not exhibit high task performance through the continuous strengthening of any single condition. Instead, the number of behavioral combinations increased substantially and became more differentiated. Compared with the low-difficulty task, Practice, Feedback, Peer Interaction, and AI Support appeared in multiple configurations in different forms, suggesting that learners may achieve similar levels of high task performance through different combinations of conditions. Previous research has suggested that this phenomenon may be related to learners' subjective perceptions of task complexity and their degree of adaptation to task demands [3]. When confronted with task requirements at a transitional level of difficulty, learners' different levels of acceptance and adaptation to changes in difficulty may lead them to select different learning resources and forms of support.

4.3.3 High-Difficulty Task

Four configurations associated with high Task Performance were identified for the high-difficulty task (see **Table 8**). The consistency values of the configurations ranged from 0.892 to 0.950, while their raw coverage values ranged from 0.327 to 0.448. **Configuration 1 (C1)** had the largest coverage among the configurations for this task. It was characterized by the peripheral absence of Practice and the core absence of both Peer Interaction and AI Support. Its raw coverage was 0.448, unique coverage was 0.123, and consistency was 0.892.

Overall, high task performance in the high-difficulty task did not show a clear dependence on Practice. Practice appeared as a peripheral absent condition in all configurations, constituting a relatively stable configurational characteristic of this task. At the same time, Engagement, Peer Interaction, Feedback, and AI Support appeared across different configurations as core present, peripheral present, or core absent conditions, forming diverse combinations. In particular, AI Support appeared as a core present condition in some configurations but as a core absent condition in others, forming different combinations with Engagement and Peer Interaction. These findings suggest that, in the high-difficulty task, high task performance was not characterized by a continuous increase in Practice, but rather by differentiated combinations of internally driven learning behaviors and external support conditions.

Table 8 Configuration for High-Difficulty Tasks (fsTask Performance)

Conditional Variables	**C1**	**C2**	**C3**	**C4**
Engagement		●		⊘
Practice	⊘	⊘	⊘	⊘
Feedback		○		⊗
Peer Interaction	⊗		●	
AI Support	⊗		●	●
Raw coverage	0.448	0.348	0.327	0.327
Unique coverage	0.123	0.023	0.055	0.039
Consistency	0.892	0.95	0.893	0.934

Note: The results of the robustness test are provided in **Appendix D.**

4.3.4 Cross-Task Comparison

A comparison of the configurational results across the three tasks reveals clear dynamic characteristics in the structure of configurations associated with high EFL task performance as task complexity changes. First, the number of configurations showed a pattern resembling an inverted U-shape as task complexity increased, increasing from a relatively small number of configurations in the low-difficulty task to a larger number in the medium-difficulty task and then decreasing in the high-difficulty task. This pattern suggests that, as task complexity increased, the combinational relationships between learning behaviors and external support underwent a process of expansion followed by further reconfiguration.

At the same time, Practice and AI Support exhibited notably different configurational patterns across tasks. As task complexity increased, the two conditions showed opposing patterns in their configurational roles. In the low-difficulty task, Practice appeared as a core present condition in **C1**, whereas AI Support was primarily characterized by its absence. In contrast, in the high-difficulty task, Practice consistently appeared as a peripheral absent condition, while AI Support appeared as either a core present or a core absent condition across different configurations.

Taken together, the results across the three tasks indicate that high task performance was not determined by any single stable behavioral condition. Rather, it was characterized by the dynamic reconfiguration of multiple learning behaviors and external support conditions in relation to task complexity.

5. Discussion

This study involved 33 participants and three GBL-based EFL tasks with different levels of difficulty, addressing RQ1 and RQ2 from a configurational perspective. The results indicate that high task performance in EFL learning was not associated with a stable and unique behavioral configuration, but could be achieved through multiple combinations of learning behaviors and support conditions. The necessity analysis further showed that no single condition reached the consistency threshold for a necessary condition for high task performance across the three tasks. Overall, the findings support the view that high task performance in GBL-based EFL learning does not follow a stable cross-task behavioral pattern but is strongly task-contingent. Task complexity may serve as an important factor driving the reconfiguration of internally driven learning behaviors and external support conditions.

The most important finding of this study is that no single condition constituted a necessary condition for high task performance across the three tasks. Although Engagement, Practice, Feedback, Peer Interaction, and AI Support all played roles in different configurations, none of these conditions consistently occupied a core position across tasks. This finding is consistent with Kolb's experiential learning theory and SRL, both of which emphasize learners' active participation and continuous self-regulation during learning [7–10]. However, this does not imply that the internally driven conditions and external scaffolding conditions conceptualized in this study were ineffective. Rather, the finding may suggest that, even in AI-enhanced learning environments, changes in task complexity can prompt the reorganization of learning resources and behaviors.

The "inverted U-shaped" pattern in the number of configurations across the three levels of task complexity further illustrates this possibility. As task complexity increased, the number of

configurations first increased and then decreased. Rather than simply interpreting this pattern as a linear relationship between task complexity and the number of learning strategies, it may be better understood as reflecting the nonlinear configuration of learning behaviors across different levels of task complexity. Notably, although the medium-difficulty task produced the largest number of high-performance configurations, the unique coverage of these configurations was relatively limited. This suggests that similar levels of high task performance could be achieved through multiple behavioral combinations, consistent with the configurational principle of equifinality, whereby different combinations of conditions can lead to similar outcomes [35].

This finding also suggests that the internally driven conditions of Engagement, Practice, and Feedback should not be understood simply as independent and universally effective behaviors. Across different levels of task complexity, these internally driven conditions may need to be combined with external scaffolding conditions in different ways to form configurations associated with high task performance. Similarly, AI Support, as a form of external scaffolding, did not appear as a condition associated with high task performance in all tasks; for example, it was primarily characterized by its absence in the relevant configurations of the low-difficulty task. This suggests that, in an era of increasingly pervasive AI-supported learning, high task performance cannot simply be equated with "more AI use," nor does it imply that learners only need to continuously increase a particular type of proactive behavior, such as greater engagement or more practice.

This task contingency was particularly evident in the cross-task changes in Practice and AI Support. Across the high-performance configurations of different tasks, these two conditions exhibited different configurational roles, including core presence and core absence. As task complexity increased, learners may have needed to reconfigure their limited time, cognitive resources, and external support. In the low-difficulty task, extensive and repeated practice may have facilitated language expression, familiarity with rules, and procedural task completion. As task complexity increased, however, learners may have needed to rely more on external scaffolding to address higher-level task demands, such as information integration, contextual decision-making, and cross-cultural communication. This interpretation is consistent with previous research on AI-supported self-regulated learning and dynamic scaffolding [10–12]. Generative AI can provide scaffolding during information acquisition, feedback, and learning-strategy adjustment, but its learning value still depends on how learners use and integrate these external resources.

However, these findings do not directly indicate that repeated practice is unimportant in complex tasks. In the high-difficulty task (T3), AI Support appeared as a core absent condition in **C1**. This indicates that some cases of high task performance may have achieved similar outcomes through other combinations of conditions in the absence of AI support. A more

precise interpretation is that, as task complexity increased, Practice and AI Support became differentiated in their configurational roles within high-performance configurations. This finding also resonates with scaffolding theory, which emphasizes the contextual dependence of scaffolding [13,15,18]. In other words, the contribution of scaffolding may depend on how well it matches the task demands currently faced by learners.

The emergence of the largest number of high-performance configurations in the medium-difficulty task also warrants further discussion. One possible explanation is that the medium-difficulty task occupied a transitional position in terms of learning demands. Previous research on task complexity has shown that learners' responses to tasks depend not only on objective task requirements but also on their own abilities and subjective perceptions of task complexity [3]. Another possibility is that the medium-difficulty task provided a more diverse space for behavioral choices, thereby allowing a greater number of identifiable combinations of conditions and, consequently, more high-performance configurations to emerge. Thus, this transitional stage does not necessarily indicate that T2 was the critical stage of the overall learning process. A more cautious interpretation is that the medium-difficulty task exhibited a particularly pronounced reconfiguration of learning behaviors.

Overall, the formation of high task performance can be understood more closely as a task-contingent process of learning resource allocation. High task performance in GBL-based EFL learning is strongly task-dependent. The key issue is not whether a particular behavior occurs "more," but whether different learning resources and actions can be effectively combined according to the specific demands of the task.

5.1 Broader implications

This study has several broader theoretical and practical implications.

At the theoretical level, unlike previous studies that have treated individual behaviors as independent predictors [e.g., 12], this study suggests that learner-driven behaviors and external support may form multiple equifinal combinations across different task-difficulty contexts. This finding contributes to a deeper understanding of GBL-based EFL learning by suggesting that learning performance cannot be explained solely in terms of whether a particular behavior is effective. Instead, attention should also be given to how different conditions jointly form configurations associated with high task performance.

This theoretical interpretation also provides implications for EFL teaching practice. In EFL instructional design, learning scaffolding may be dynamically adjusted according to task complexity rather than providing exactly the same type and level of support across all tasks. In particular, educators should pay attention to how both teachers and students perceive and use AI. Rather than treating AI as a fixed learning tool that is equally applicable to all tasks, AI can

be conceptualized as an external learning resource whose level and form of support can be flexibly adjusted according to task requirements.

This study also has practical implications for vocational English education in higher education. Professional English instruction and project-based learning designs may benefit from a more explicit progression of learning contexts, with task demands increasing progressively throughout the learning process. Such an approach can provide learners with greater space for autonomous engagement while allowing teachers to dynamically adjust the intensity of peer interaction and AI support according to the progression of task requirements, thereby creating a learning support environment that is better aligned with different task demands.

5.2 Limitations

This study has several limitations that should be considered in future research.

First, the sample size and source limit the cross-contextual stability of the findings. The sample was drawn from 33 students majoring in Tourism Management at a single university. Although each participant completed three tasks, resulting in 99 task-level observations, the number of cases in the fsQCA remained 33 participants. Future research could therefore expand the sources of participants by including students from different universities, disciplinary backgrounds, and levels of study, thereby examining the cross-contextual applicability of the findings.

Second, the classification of task complexity involves a certain degree of uncertainty. Although all three tasks were conducted within an EFL course, they differed substantially in their topics and content. Such differences may have affected comparisons of task complexity across the three tasks. Future research could maintain greater consistency in task topics and content while systematically manipulating task complexity to further examine how learning-behavior configurations change with task complexity.

Third, the behavioral measurement approach adopted in this study could not fully capture learners' cognitive and psychological processes. Although questionnaire-based measures may be subject to subjective bias, they can effectively capture learners' attitudes, perceptions, and psychological states. Future research could therefore combine behavioral trace data with interviews and process-based questionnaires to incorporate both objective behavioral evidence and learners' subjective cognitive and psychological experiences, thereby providing a more comprehensive understanding of learning behaviors and their underlying mechanisms.

Finally, this study shares a limitation that is relatively common in fsQCA research [e.g., 36,37], namely that configurational analysis does not provide a sufficiently fine-grained longitudinal account of how individual cases evolve across different configurations. In other

words, the present study identifies the combinations of conditions associated with high task performance under different task contexts, but it cannot further determine whether individual learners consistently adopt relatively stable behavioral strategies or how their strategies are adjusted as task complexity changes. Future research could therefore incorporate case tracking and longitudinal process data to continuously examine individual learners, thereby exploring the stability and adjustment of their learning strategies and their relationship with changes in task context.

6. Conclusion

This study employed fsQCA to examine learning-behavior configurations in GBL-based EFL learning across tasks with different levels of complexity, considering both internally driven behaviors and external support conditions. The findings indicate that high task performance does not follow a single and stable pattern but can emerge from different combinations of internal and external conditions. The variation in configurational structures across tasks of different difficulty further demonstrates the dynamic nature of learning-behavior configurations, particularly in the differentiated configurational roles of Practice and AI Support across tasks. By adopting a configurational perspective, this study provides further empirical evidence for understanding learning behaviors and their task-contingent nature in complex learning environments. Based on these findings, educators should avoid treating AI and other forms of external scaffolding as universally applicable forms of support. Instead, learning environments should be designed with greater flexibility and contextual adaptability according to specific task requirements and learning contexts.

7. Data Availability Statement

The data used in this study are managed by the Academic Affairs Office and the Research Office of Wuhan Institute of Technology and Business. The data are available from the corresponding author upon reasonable request, subject to approval by the relevant institutional departments.

Reference

1. Almayez, M., Al-khresheh, M. H., Alruwaili, S. F., Al-Ali, M. N. & Alenezy, F. S. AI-driven adaptive feedback and EFL writing performance: The roles of engagement, metacognition, and epistemic agency in a cross-linguistic context. *Acta Psychologica* **268**, 107322 (2026).
2. Zhou, N. Disciplinary English reform and practices of undergraduate program in response to cultivation of interdisciplinary elites in China. *Humanit Soc Sci Commun* **12**, 1140 (2025).
3. Hamari, J. *et al.* Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior* **54**, 170–179 (2016).
4. Jiang, W., Li, D. & Liu, C. Understanding dimensions of trust in AI through quantitative cognition: Implications for human-AI collaboration. *PLoS One* **20**, e0326558 (2025).
5. Herrmann-Werner, A. *et al.* Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study. *J Med Internet Res* **26**, e52113 (2024).
6. Faza, A. & Lestari, I. A. Self-Regulated Learning in the Digital Age: A Systematic Review of Strategies, Technologies, Benefits, and Challenges. *IRRODL* **26**, 23–58 (2025).
7. Hung, L.-Y., Wang, S.-M. & Yeh, T.-K. Kolb's experiential learning theory and marine debris education: Effects of different stages on learning. *Marine Pollution Bulletin* **191**, 114933 (2023).
8. Kim, M. & Park, M. J. Absorptive capacity in entrepreneurial education: Rethinking the Kolb's experiential learning theory. *The International Journal of Management Education* **21**, 100873 (2023).
9. Zimmerman, B. J. Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist* **25**, 3–17 (1990).
10. Xia, Q., Liu, Q., Tlili, A. & Chiu, T. K. F. A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education* **240**, 105465 (2026).
11. Xu, X., Qiao, L., Cheng, N., Liu, H. & Zhao, W. Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *Brit J Educational Tech* **56**, 1842–1863 (2025).
12. Lan, M. & Zhou, X. A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Sci. Learn.* **10**, 21 (2025).
13. Wilson, K. & Devereux, L. Scaffolding theory: High challenge, high support in Academic Language and Learning (ALL) contexts. *Journal of Academic Language and Learning* **8**, A91–A100 (2014).
14. Xi, J. & Lantolf, J. P. Scaffolding and the zone of proximal development: A problematic relationship. *J Theory Soc Behav* **51**, 25–48 (2021).
15. Hammond, J. High challenge, high support: Integrating language and content instruction for diverse learners in an English literature classroom. *Journal of English for Academic Purposes* **5**, 269–283 (2006).
16. Jiang, W., Li, Y., Hu, X. & Ma, D. Reframing individual roles in collaboration: digital

identity construction and adaptive mechanisms for resistance-based professional skills in AI-human intelligence symbiosis. *Front. Psychol.* **16**, 1652130 (2025).
17. Shao, X., Jiang, W. & Osman, K. N. Beyond facilitation and inhibition: a configurational mechanism study of cognitive transitions in human–AI collaboration. *Front. Psychol.* **17**, 1821188 (2026).
18. Siu, K. W. M. *et al.* Dynamic scaffolding: exploring the role of artificial intelligence in urban design education. *Front. Urban Rural Plan.* **3**, 8 (2025).
19. Yang, W., Lu, X. & Weigle, S. C. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing* **28**, 53–67 (2015).
20. Yang, Q.-F., Chang, S.-C., Hwang, G.-J. & Zou, D. Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. *Computers & Education* **148**, 103808 (2020).
21. Du, J., Hew, K. F. & Liu, L. What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. *Computers & Education* **201**, 104828 (2023).
22. Alhazbi, S. *et al.* Using learning analytics to measure SELF-REGULATED learning: A systematic review of empirical studies in higher education. *Computer Assisted Learning* **40**, 1658–1674 (2024).
23. Ye, D. & Pennisi, S. Using trace data to enhance Students' self-regulation: A learning analytics perspective. *The Internet and Higher Education* **54**, 100855 (2022).
24. Rihoux, B. & Ragin, C. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. (SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 2009). doi:10.4135/9781452226569.
25. Boulahmel, A., Djelil, F. & Smits, G. Investigating Self-Regulated Learning Measurement Based on Trace Data: A Systematic Literature Review. *Tech Know Learn* **30**, 119–156 (2025).
26. Banihashem, S. K., Noroozi, O., Van Ginkel, S., Macfadyen, L. P. & Biemans, H. J. A. A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review* **37**, 100489 (2022).
27. Soda, G. & Zaheer, A. A network perspective on organizational architecture: performance effects of the interplay of formal and informal organization. *Strategic Management Journal* **33**, 751–771 (2012).
28. Saqr, M., Elmoazen, R., Tedre, M., López-Pernas, S. & Hirsto, L. How well centrality measures capture student achievement in computer-supported collaborative learning? – A systematic review and meta-analysis. *Educational Research Review* **35**, 100437 (2022).
29. Carpenter, S. K. Testing Enhances the Transfer of Learning. *Curr Dir Psychol Sci* **21**, 279–283 (2012).
30. Holdsworth, S., Sandri, O., Hayes, J. & Maslen, S. Scenario-based capability assessment in professional learning contexts. *Journal of Adult and Continuing Education* **31**, 600–627 (2025).
31. Jiang, W., Ma, D. & Yang, H. Research on the Optimization of Soil and Water

Conservation Schemes Based on GenAI: A Hybrid Approach Using fsQCA and Machine Learning. in *13th International Symposium on Project Management, ISPM 2025, June 27, 2025 - June 28, 2025* vol. 2 662–667 (Aussino Academic Publishing House, Beijing, China, 2025).

32. Sinaga, K. P. & Yang, M.-S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **8**, 80716–80727 (2020).
33. Dong, H., Chen, X. & Shen, J. Unpacking the impact mechanism of generative AI-based learning analytics technology on student learning engagement: the mediating role of community of inquiry and the moderating role of digital literacy. *Int J Educ Technol High Educ* **23**, 14 (2026).
34. Zhao, Y. & Zhang, J. Beyond Using AI : Self-Regulated Pathways to Engagement in AI - Assisted Language Learning. *Euro J of Education* **61**, e70842 (2026).
35. Pappas, I. O. & Woodside, A. G. Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. *International Journal of Information Management* **58**, 102310 (2021).
36. Pappas, I. O., Kourouthanassis, P. E., Giannakos, M. N. & Chrissikopoulos, V. Explaining online shopping behavior with fsQCA: The role of cognitive and affective perceptions. *Journal of Business Research* **69**, 794–803 (2016).
37. Feng, N., Zhao, W. & Wei, F. The Willingness of SMEs to Participate in Key Core Technology Breakthroughs under the Organizational Model of an Innovation Ecosystem——An Integrated Approach Using PLS-SEM and fsQCA. *Science & Technology Progress and Policy* 1–12 (2026) doi:https://doi.org/10.6049/kjjbydc.D2025090586.

Appendix A.

Table A1. Measurement and Standardization of Task Performance

Measurement dimension	Post-test	Scenario-based Transfer Report
Purpose	To assess learners' mastery of the target knowledge associated with the task	To assess learners' ability to transfer acquired knowledge to a new context and apply it comprehensively
T1 Food and Beverage	Food and beverage vocabulary, ordering expressions, and language for describing and recommending dishes	Information expression, understanding of customer needs, and dish recommendation in a new food-and-beverage context
T2 Shopping for Souvenirs	Shopping vocabulary, product description and comparison, and price-negotiation expressions	Product comparison, selection, description, and bargaining in a new shopping context
T3 Sightseeing	Language knowledge related to attraction descriptions, tourism information, and routes, time, and budget	Destination planning, integration of routes and budget, attraction explanation, and contextualized expression in a new tourism context
Scoring	Scored according to a predefined scoring rubric	Scored according to a predefined rubric for the scenario-based task
Weight	40%	60%
Standardization	Raw scores were converted to a 0 – 100 scale	Raw scores were converted to a 0 – 100 scale
Final score	Task Performance = Post-test × 0.40 + Transfer report × 0.60	

Table A2. Necessity Analysis of ~fsTask Performance (Low Task Performance)

Conditions	Task1		Task2		Task3	
	Consistency	Coverage	Consistency	Coverage	Consistency	Coverage
fsEngagement	0.356834	0.735969	0.502582	0.333295	0.456816	0.393369
~fsEngagement	0.87817	0.564388	0.929432	0.222864	0.947971	0.417125
fsPractice	0.684601	0.726855	0.435456	0.168892	0.258064	0.48723
~fsPractice	0.586271	0.533483	0.943201	0.304107	0.936524	0.322465
fsFeedback	0.777366	0.624752	0.75043	0.257531	0.774193	0.395324
~fsFeedback	0.497217	0.624224	0.562823	0.203485	0.642039	0.43512
fsPeer Interaction	0.642548	0.613704	0.707401	0.260952	0.676379	0.410095
~fsPeer Interaction	0.603587	0.607343	0.681583	0.229565	0.585848	0.32828
fsAI Support	0.611627	0.651086	0.743545	0.282538	0.62539	0.44387
~fsAI Support	0.642548	0.58338	0.654045	0.214568	0.815817	0.402878

Appendix B. Robustness Tests for Task 1

Table B1. Configurational Analysis Results for Task 1 at a Consistency Threshold of 0.80

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
Complex solution	Engagement * Practice * Feedback * PeerInteraction	0.251	0.15	0.802
	~Engagement * ~Practice * Feedback * ~Peer Interaction * ~AI Support	0.359	0.201	0.843
	~Engagement * ~Practice * ~Feedback * Peer Interaction * ~AI Support	0.248	0.102	0.812
Solution	—	0.618	—	0.791
Parsimonious solution	~Practice * Feedback * ~Peer Interaction * ~AI Support	0.359	0.149	0.843
	~Practice * ~Feedback * Peer Interaction * ~AI Support	0.262	0.103	0.82
	Engagement * Feedback	0.318	0.009	0.725
	Engagement * Practice * Peer Interaction	0.259	0.008	0.777
Solution	—	0.637	—	0.756
Intermediate solution	Engagement * Practice * Feedback * Peer Interaction	0.251	0.15	0.802
	~Engagement * ~Practice * Feedback * ~Peer Interaction * ~AI Support	0.359	0.201	0.843
	~Engagement * ~Practice * ~Feedback * Peer Interaction * ~AI Support	0.248	0.102	0.812
Solution	—	0.618	—	0.791

Note. Frequency threshold = 1; consistency threshold = 0.8. "~" denotes the absence of a condition.

Table B2. Configurational Analysis Results for Task 1 at a Consistency Threshold of 0.78

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
Complex solution	~Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.446	0.286	0.799
	~Engagement * ~Feedback * Peer Interaction * ~AI Support	0.272	0.117	0.78
	Engagement * Practice * Feedback * Peer Interaction	0.251	0.143	0.802
Solution	—	0.72	—	0.761
Parsimonious solution	~Practice * ~Peer Interaction * ~AI Support	0.446	0.234	0.799
	~Feedback * Peer Interaction * ~AI Support	0.288	0.114	0.79
	Engagement * Feedback	0.318	0.009	0.725
	Engagement * Practice * Peer Interaction	0.259	0.004	0.777
Solution	—	0.735	—	0.731
Intermediate solution	~Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.446	0.286	0.799
	~Engagement * ~Feedback * Peer Interaction * ~AI Support	0.272	0.117	0.78
	Engagement * Practice * Feedback * Peer Interaction	0.251	0.143	0.802
Solution	—	0.72	—	0.761

Note. Frequency threshold = 1; consistency threshold = 0.78. "~" denotes the absence of a condition.

Table B3. Parameter Variations in the Robustness Test for Task 1

Consistency threshold	Solution type	Configurations	Coverage	Consistency
0.8	Complex	3	0.618	0.791
0.8	Intermediate	3	0.618	0.791
0.8	Parsimonious	4	0.637	0.756
0.78	Complex	3	0.72	0.761
0.78	Intermediate	3	0.72	0.761
0.78	Parsimonious	4	0.735	0.73

Note: As shown in Appendix B, the configuration structures identified for Task 1 remained generally stable when the consistency threshold was changed from 0.80 to 0.78, although solution coverage and consistency varied to some extent.

Appendix C. Robustness Tests for Task 2

Table C2. Configurational Analysis Results for Task 2 at a Consistency Threshold of 0.89

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
Complex solution	~Engagement * Practice	0.444	0.074	0.98
	~Practice * Feedback * ~AI Support	0.278	0.048	0.918
	~Engagement * Feedback * Peer Peer Interaction	0.323	0.029	0.952
	Practice * ~Peer Interaction * AI Support	0.256	0.043	0.996
	Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.149	0.038	0.973
	~Engagement * ~Feedback * ~Peer Interaction * AI Support	0.197	0.047	0.973
	Practice * ~Feedback * Peer Interaction * AI Support	0.192	0.023	0.994
Solution	—	0.785	—	0.956
Parsimonious solution	~AI Support	0.597	0.115	0.916
	Practice	0.539	0.1	0.978
	~Feedback * ~Peer Interaction	0.367	0.049	0.961
	Feedback * Peer Interaction	0.334	0.028	0.902
Solution	—	0.903	—	0.917
Intermediate solution	~Engagement * Practice	0.444	0.074	0.98
	~Practice * Feedback * ~AI Support	0.278	0.048	0.918
	~Engagement * Feedback * Peer Interaction	0.323	0.029	0.952
	Practice * ~Peer Interaction * AI Support	0.256	0.043	0.996
	Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.149	0.038	0.973
	~Engagement * ~Feedback * ~Peer Interaction * AI Support	0.197	0.047	0.973
	Practice * ~Feedback * Peer Interaction * AI Support	0.192	0.023	0.994
Solution	—	0.785	—	0.956

Note. Frequency threshold = 1; consistency threshold = 0.89. "~" denotes the absence of a condition.

Table C2. Configurational Analysis Results for Task 2 at a Consistency Threshold of 0.9

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
	~Engagement * Practice	0.444	0.128	0.98

	Practice * ~Peer Interaction * AI Support	0.256		0.043	0.996
	Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.149		0.063	0.973
Complex solution	~Engagement * ~Feedback * ~Peer Interaction * AI Support	0.197		0.049	0.973
	Practice * ~Feedback * Peer Interaction * AI Support	0.192		0.023	0.994
	~Practice * Feedback * ~Peer Interaction * ~AI Support	0.204		0.071	0.922
Solution	—	0.715	—		0.963
	Engagement	0.307		0.019	0.953
Parsimonious	Practice	0.539		0.13	0.978
solution	~Feedback * ~Peer Interaction	0.367		0.068	0.961
	Peer Interaction * ~AI Support	0.371		0.087	0.92
Solution	—	0.805	—		0.936
	~Engagement * Practice	0.444		0.128	0.98
	Practice * ~Peer Interaction * AI Support	0.256		0.043	0.996
	Engagement * ~Practice * ~Peer Interaction * ~AI Support	0.149		0.063	0.973
Intermediate solution	~Engagement * ~Feedback * ~Peer Interaction * AI Support	0.197		0.049	0.973
	Practice * ~Feedback * Peer Interaction * AI Support	0.192		0.023	0.994
	~Practice * Feedback * ~Peer Interaction * ~AI Support	0.204		0.071	0.922
Solution	—	0.715	—		0.963

Note. Frequency threshold = 1; consistency threshold = 0.90. "~" denotes the absence of a condition.

Table C3. Parameter Variations in the Robustness Test for Task 2

Consistency threshold	Solution type	Configurations	Coverage	Consistency
0.89	Complex	7	0.785	0.956
0.89	Intermediate	7	0.785	0.956
0.89	Parsimonious	4	0.903	0.917
0.9	Complex	6	0.715	0.963
0.9	Intermediate	6	0.715	0.963
0.9	Parsimonious	4	0.805	0.936

Note: As shown in Appendix C the configuration structures identified for Task 2 remained generally stable when the consistency threshold was changed from 0.89 to 0.90, although solution coverage and consistency varied to some extent.

Appendix D. Robustness Tests for Task 3

Table D1. Configurational Analysis Results for Task 3 at a Consistency Threshold of 0.91

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
Complex solution	~Practice * ~Peer Interaction * ~AI Support	0.448	0.076	0.892
	Engagement * ~Practice * Feedback	0.348	0.023	0.95
	~Practice * Peer Interaction * AI Support	0.327	0.134	0.893
	~Engagement * ~Practice * ~Feedback * ~Peer Interaction	0.298	0.05	0.931
Solution	—	0.71	—	0.877
Parsimonious solution	Engagement	0.456	0.06	0.955
	~Peer Interaction * ~AI Support	0.476	0.071	0.888
	Peer Interaction * AI Support	0.342	0.044	0.895
	~Feedback * ~Peer Interaction	0.35	0.01	0.935
	~Feedback * AI Support	0.377	0	0.942
Solution	—	0.79	—	0.88
Intermediate solution	~Practice * ~Peer Interaction * ~AI Support	0.448	0.123	0.892
	Engagement * ~Practice * Feedback	0.348	0.023	0.95
	~Practice * Peer Interaction * AI Support	0.327	0.055	0.893
	~Engagement * ~Practice * ~Feedback * AI Support	0.327	0.039	0.934
Solution	—	0.699	—	0.883

Note. Frequency threshold = 1; consistency threshold = 0.91. "~" denotes the absence of a condition.

Table D2. Configurational Analysis Results for Task 3 at a Consistency Threshold of 0.89

Solution type	Configuration	Raw coverage	Unique coverage	Consistency
Complex solution	~Practice * ~Peer Interaction * ~AI Support	0.448	0.018	0.892
	Engagement * ~Practice * Feedback	0.348	0.006	0.95

	~Practice * Peer Interaction * AI Support	0.327		0.03	0.893
	~Practice * ~Feedback * ~AI Support	0.537		0.016	0.901
	~Practice * ~Feedback * Peer Interaction	0.382		0.006	0.87
	~Engagement * ~Practice * ~Feedback * ~Peer Interaction	0.298		0.01	0.931
	~Engagement * ~Practice * ~Feedback * AI Support	0.327		0	0.934
Solution	—	0.79	—		0.87
Parsimonious solution	Engagement	0.456		0.043	0.955
	~Peer Interaction * ~AI Support	0.476		0.013	0.888
	~Feedback * ~Peer Interaction	0.35		0.01	0.935
	Feedback * ~AI Support	0.55		0.023	0.88
	~Feedback * AI Support	0.377		0	0.942
	Feedback * Peer Interaction	0.388		0.012	0.871
	Peer Interaction * AI Support	0.342		0.019	0.895
Solution	—	0.884	—		0.863
Intermediate solution	~Practice * ~Peer Interaction * ~AI Support	0.448		0.018	0.892
	Engagement * ~Practice * Feedback	0.348		0.006	0.95
	~Practice * Peer Interaction * AI Support	0.327		0.03	0.893
	~Practice * ~Feedback * ~AI Support	0.537		0.016	0.901
	~Practice * ~Feedback * Peer Interaction	0.382		0.006	0.87
	~Engagement * ~Practice * ~Feedback * ~Peer Interaction	0.298		0.01	0.931
	~Engagement * ~Practice * ~Feedback * AI Support	0.327		0	0.934
Solution	—	0.79	—		0.87

Note. Frequency threshold = 1; consistency threshold = 0.89. "~" denotes the absence of a condition.

Table D3. Parameter Variations in the Robustness Test for Task 3

Consistency threshold	Solution type	Configurations	Coverage	Consistency
0.91	Complex	4	0.71	0.877
0.91	Intermediate	4	0.699	0.883
0.91	Parsimonious	5	0.79	0.88
0.89	Complex	7	0.79	0.87
0.89	Intermediate	7	0.79	0.87
0.89	Parsimonious	7	0.884	0.863

Note: As shown in Appendix D, the configuration structures identified for Task 3 remained generally stable when the consistency threshold was changed from 0.89 to 0.91, although solution coverage and consistency varied to some extent.