Asian Journal of Education and Training
Vol. 11, No. 4, 224-232, 2025
ISSN(E) 2519-5387
DOI: 10.20448/edu.v11i4.7943

AIGC-assisted evaluation of teachers' tour-guide scripts: Construction and practice of the six-step method

Lingman Zhou[1]
Ting Xu[2]
Yunyao Zhang[3]✉

(✉ *Corresponding Author)*

[1,2,3]*Department of Geography and Tourism, Chongqing Normal University, Chongqing, China.*
[1]*Email: 18875032334@163.com*
[2]*Email: 13684109025@163.com*
[3]*Email: zhyunyao@126.com*

Abstract

Artificial Intelligence Generated Content (AIGC) has garnered significant attention in education due to its strengths in language processing and content creation. It offers a novel technical approach to enhance tools for evaluating tour-guide scripts, which are inherently contextual, audience-oriented, and focused on cultural expression. This paper proposes a six-step method for AIGC-assisted evaluation of teachers' tour-guide scripts within the framework of Technological Pedagogical Content Knowledge (TPACK). The approach leverages the integration of technological knowledge, pedagogical knowledge, and content knowledge to establish a clear human–AI workflow. The six steps include Standard Setting, Standardized Input, AIGC Check, Guided Modification, Comparative Analysis, and Summary and Sharing. The study utilizes vocational students as the research group and employs a combination of teacher–AIGC comparison and mixed methods for evaluation. The focus is on the quantity and types of comments, the focus dimensions, and students' performance in second-round writing. Results indicate that AIGC is effective in ensuring language accuracy and structural completeness. Teachers demonstrate stronger professional judgment in areas related to context and cultural expression. Their collaboration within the six-step process significantly enhances evaluation efficiency. This method provides a clear, repeatable process model and offers practical support for integrating AIGC into tour-guide script teaching.

Keywords: AIGC, TPACK, Tour-guide scripts, Writing assessment, Human-AI collaboration, Secondary vocational education.

Citation | Zhou, L., Xu, T., & Zhang, Y. (2025). AIGC-assisted evaluation of teachers' tour-guide scripts: Construction and practice of the six-step method. *Asian Journal of Education and Training, 11*(4), 224–232. https://doi.org/10.20448/edu.v11i4.7943
History:
Received: 8 July 2025
Revised: 10 December 2025
Accepted: 18 December 2025
Published: 24 December 2025

Publisher: Asian Online Journal Publishing Group

Funding: This work was supported by Chongqing Normal University of China. (Grant number: 202531).

Institutional Review Board Statement: This study was approved by the Institutional Review Board of Chongqing Normal University, China, under protocol number CQUDLLY070501, dated July 5, 2025. Informed verbal consent was obtained from all participants, and all data were anonymized to protect participant confidentiality.

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.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

Disclosure of AI Use: The authors used OpenAI's ChatGPT to edit and refine the wording of the text. All outputs were reviewed and verified by the authors. Please see the revised manuscript.

Contents

Contribution of this paper to the literature

This paper proposes the SICMCS six-step method within the TPACK framework for evaluating tour-guide scripts. This method standardizes the process and clarifies the roles of humans and AI in a closed classroom loop. By comparing teacher feedback with AIGC feedback, the study suggests adopting a teacher-led approach assisted by AIGC.

1. Introduction

With the widespread use of AI in education and strong policy support, educational assessment is undergoing a system-wide redesign. The 2020 Overall Plan for Advancing Educational Evaluation Reform in the New Era emphasizes the need for developing new assessment methods. It advocates for the use of AI and big data to enhance the scientific validity, professional quality, and objectivity of assessments (China, 2020). Driven by technological advancements and policy initiatives, AI-based writing assessment is expanding. However, most current tools are limited to surface features such as vocabulary and grammar. They do not support the diagnosis and improvement of higher-order writing skills (Cao, 2022; Li, 2020). AIGC, represented by ChatGPT, demonstrates significant potential for generative dialogue and process feedback. It can complement traditional tools by providing revision suggestions, multi-dimensional evaluations, and personalized guidance (Li, Link, & Hegelheimer, 2015; Liu, Li, & Yang, 2021). This paper uses TPACK as the theoretical foundation. It proposes the SICMCS six-step method, which integrates technological ability (T), pedagogical strategies (P), and subject content for tour-guide scripts (C). The method forms a closed-loop assessment process involving process diagnosis, targeted revision, and transfer of learning. The study utilizes tour-guide scripts from secondary vocational students and compares teacher evaluations with AIGC evaluations based on SICMCS. It combines quantitative coding with qualitative case studies to test effects in real classroom settings, focusing on revised writing quality, feedback specificity, and teacher workload. The research assesses the effectiveness and feasibility of the SICMCS process for evaluating tour-guide scripts with AIGC. It clarifies AIGC's advantages, limitations, and optimal collaboration roles, providing a repeatable framework and prompt templates that offer both theoretical insights and practical guidance for teachers using AIGC in guide script evaluation.

2. Related Theory and Research

2.1. Research on the Effects of Intelligent Technology-Enhanced Writing Assessment

Work on writing assessment with intelligent technology can be traced to Page's PEG system (Page, 1967), which pioneered the use of machine scoring. As technology has advanced, Automated Writing Evaluation (AWE) has moved into classroom practice, especially in second-language writing (Roscoe, Wilson, Johnson, & Mayra, 2017). Research shows that AWE helps improve students' writing motivation and self-efficacy (Nazari, Shabbir, & Setiawan, 2021). However, its feedback mostly targets surface features such as vocabulary, grammar, and spelling. It is limited in diagnosing higher-order skills such as content, purpose, semantic coherence, and discourse strategies. In Chinese writing, AWE improves performance and the writing process by strengthening evidence, explanation, and conclusions, and it promotes persistence and positive beliefs about revision.

AIGC has grown rapidly in recent years. Its leading product, ChatGPT, is more advanced than traditional systems in coherent expression, maintaining context, and providing interactive feedback. It can perform quality checks, as well as summarization, dialogue guidance, and targeted revisions (Chen & Cheng, 2008; Deng & Lin, 2023). It has been shown to support tiered assessment and guidance, from basic error correction and mid-level polishing to advanced rewriting (Wei & Li, 2023). Tour-guide scripts are task-based, highly contextual, and emphasize audience awareness and cultural expression. AIGC's generative and interactive strengths enable it to go beyond surface-level language correction, supporting high-level content organization and context-aware communication strategies. However, the effectiveness of AIGC in classroom settings, its interpretability, and its boundaries for working with teachers' professional judgment still require more context-rich evidence. Based on this, the paper proposes the SICMCS six-step method within the TPACK framework. This method aims to establish a process improvement loop, encompassing diagnosis, revision, and transfer of learning.

2.2. A Comparative Study of Intelligent Writing Assessment and Teacher Writing Assessment

Existing studies that compare intelligent evaluation tools with teacher evaluation mostly focus on evaluation types, evaluation levels, and accuracy (Wei & Li, 2023). By evaluation type, we usually distinguish direct and indirect evaluation. Direct evaluation gives fixes or makes edits. Indirect evaluation only marks problems and leaves revision to the student (Ellis, 2009). Evidence shows that ChatGPT tends to provide more extensive direct evaluations, while teachers prefer indirect evaluations to promote students' thinking (Guo & Wang, 2024). At the lower levels of evaluation, the focus is on spelling, punctuation, and grammar, while the higher-level evaluation concentrates on content choice, structure, main ideas, and audience strategies (AbuSeileek, 2013). Link, Mehrzad, and Rahimi (2022) found that automated writing evaluation (AWE) provides limited support for higher-level evaluation by teachers, whereas teachers not using AWE tend to give more comprehensive lower-level feedback. For evaluation accuracy, information retrieval and machine learning often utilize metrics such as precision and recall; higher values indicate better evaluation quality (Hoang & Kunnan, 2016). Empirical comparisons demonstrate that intelligent systems and teachers have similar precision levels. Although the system's recall is sometimes higher, it still requires further improvement (Liu et al., 2021).

2.3. Related Research on TPACK Theory

TPACK emphasizes integrating technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK) (Chi & Shi, 2012). It also utilizes interaction domains such as PCK (pedagogy and content knowledge), TPK (technology and pedagogy knowledge), and TCK (technology and content knowledge), as well as TPACK itself, to guide improvements in classroom practice (Huang, 2019). Unlike general writing, tour-guide scripts constitute a genre characterized by strong contextual, audience, and cultural features. They require cultural accuracy and careful information selection at the content level, emphasizing task-driven design and formative assessment. Additionally, they benefit from Artificial Intelligence Generated Content (AIGC) to provide efficient, interpretable, and transferable process support. Existing research has validated the effectiveness of TPACK in subject integration, teacher training, and curriculum development. However, there remains a lack of practical models tailored for genre-based writing, such as tour-guide scripts, and for human-AI collaboration in the evaluation process. To address this gap, this paper anchors TPACK specifically to the SICMCS six-step method, facilitating the transition from theoretical framework to an executable process.

3. The Construction and Practical Validation of the SICMCS Six-Step Method

3.1. The Construction of the SICMCS Six-Step Method

The TPACK framework is a teacher knowledge integration model proposed by American scholars Koehler and Mishra (2005) (as shown in Figure 1).

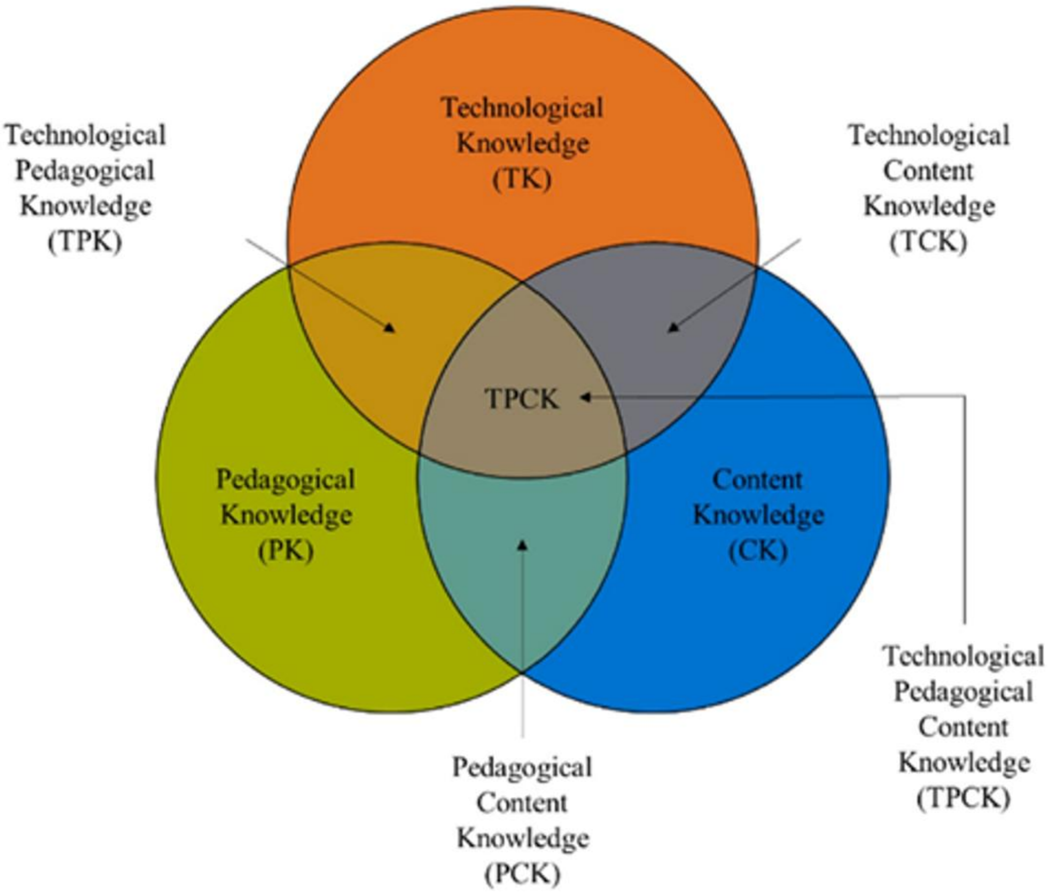

Technological Pedagogical Knowledge (TPK)
Technological Knowledge (TK)
Technological Content Knowledge (TCK)
TPCK
Pedagogical Knowledge (PK)
Content Knowledge (CK)
Technological Pedagogical Content Knowledge (TPCK)
Pedagogical Content Knowledge (PCK)

Figure 1. The TPACK framework.

The framework emphasizes the convergence of three key domains: content knowledge (CK), pedagogical knowledge (PK), and technological knowledge (TK). This intersection leads to the development of pedagogical content knowledge (PCK), technological pedagogical knowledge (TPK), and technological content knowledge (TCK), ultimately integrating all three into TPACK to guide teachers in delivering effective instruction within specific contexts (Liu, 2019). In the TPACK framework, a teacher's ability to assess effectively is developed through the integration of three key areas of knowledge: technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK). TPACK emphasizes using the task context as a platform, integrating technological tools and teaching strategies into specific subject content, thereby creating a feasible and reusable teaching assessment process. For application writing, such as tour guide scripts, which are task-based and highly contextualized, this study proposes the SICMCS six-step method (as shown in Figure 2), converting theoretical support into a practical model.

The six-step method of SICMCS comprises six stages: setting standards (S), standardized input (I), AIGC Check (C), guided modification (M), comparative analysis (C), and summary sharing (S). Based on the TPACK theory, it forms a closed-loop model consisting of standards, inputs, diagnosis, revision, comparison, and summary. CK defines what needs to be evaluated, PK determines how to teach and assess, and TK ensures implementation and automation. The collaboration of these three elements has resulted in a replicable assessment process and toolkit. Compared with traditional artificial intelligence, the capabilities of AIGC in semantic understanding and interactive generation offer technical feasibility and efficiency advantages for process diagnosis and targeted revisions (Zhang & Wang, 2024).

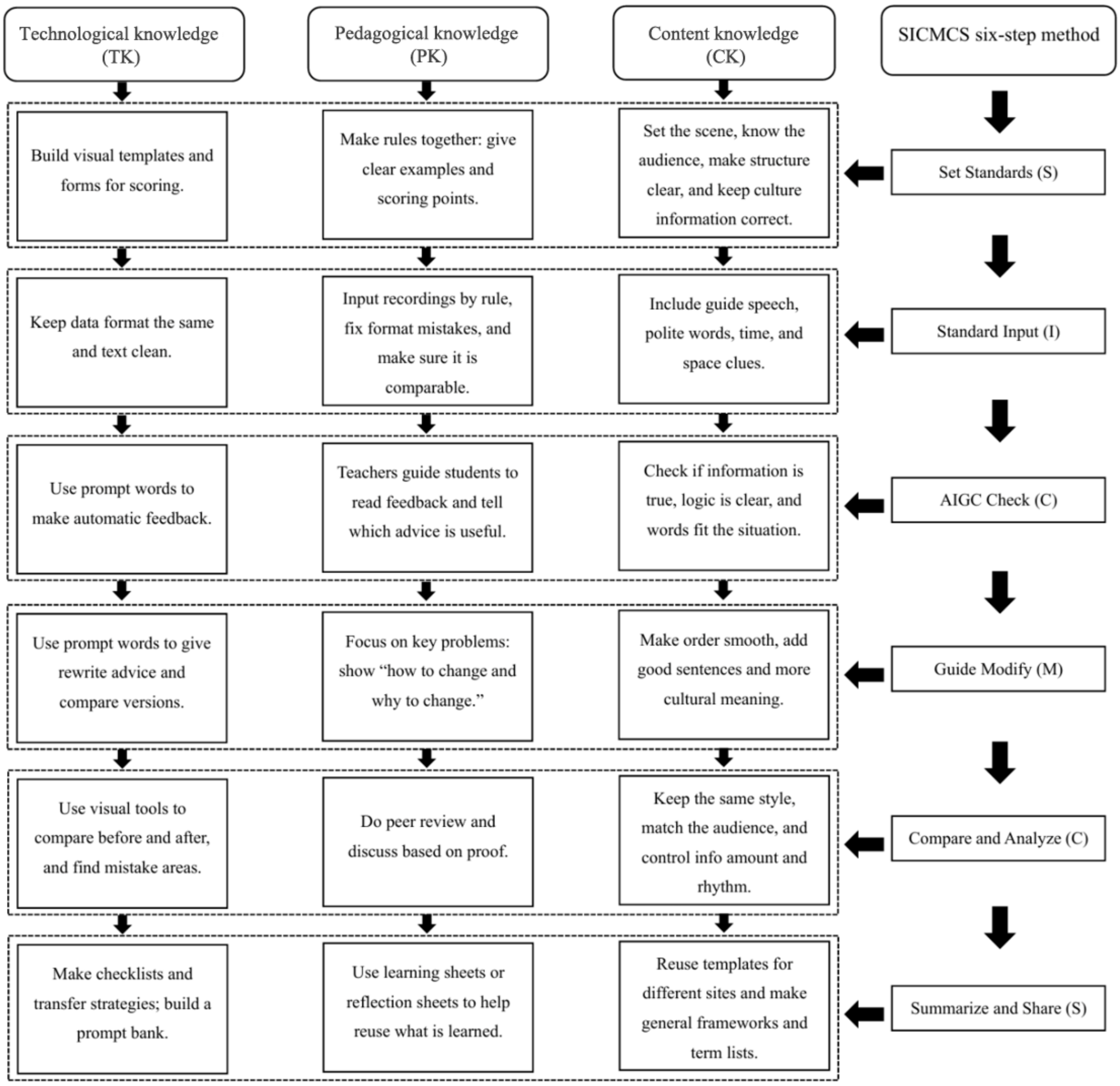

Technological knowledge (TK)
Pedagogical knowledge (PK)
Content knowledge (CK)
SICMCS six-step method
Build visual templates and forms for scoring.
Make rules together: give clear examples and scoring points.
Set the scene, know the audience, make structure clear, and keep culture information correct.
Set Standards (S)
Keep data format the same and text clean.
Input recordings by rule, fix format mistakes, and make sure it is comparable.
Include guide speech, polite words, time, and space clues.
Standard Input (I)
Use prompt words to make automatic feedback.
Teachers guide students to read feedback and tell which advice is useful.
Check if information is true, logic is clear, and words fit the situation.
AIGC Check (C)
Use prompt words to give rewrite advice and compare versions.
Focus on key problems: show "how to change and why to change."
Make order smooth, add good sentences and more cultural meaning.
Guide Modify (M)
Use visual tools to compare before and after, and find mistake areas.
Do peer review and discuss based on proof.
Keep the same style, match the audience, and control info amount and rhythm.
Compare and Analyze (C)
Make checklists and transfer strategies; build a prompt bank.
Use learning sheets or reflection sheets to help reuse what is learned.
Reuse templates for different sites and make general frameworks and term lists.
Summarize and Share (S)

Figure 2. The "SICMCS six-step method" under the TPACK framework.

3.2. Practical Validation Design

This study employs a tour-guide script evaluation lesson as a practical scenario, introducing and implementing the SICMCS six-step method within the TPACK framework. By establishing six steps setting standards, standardized input, intelligent correction, guided revision, comparative analysis, and summary sharing a classroom evaluation loop is created. This approach provides a unified research framework and operational pathway for comparing teacher evaluations with AIGC evaluations.

3.2.1. Research Object

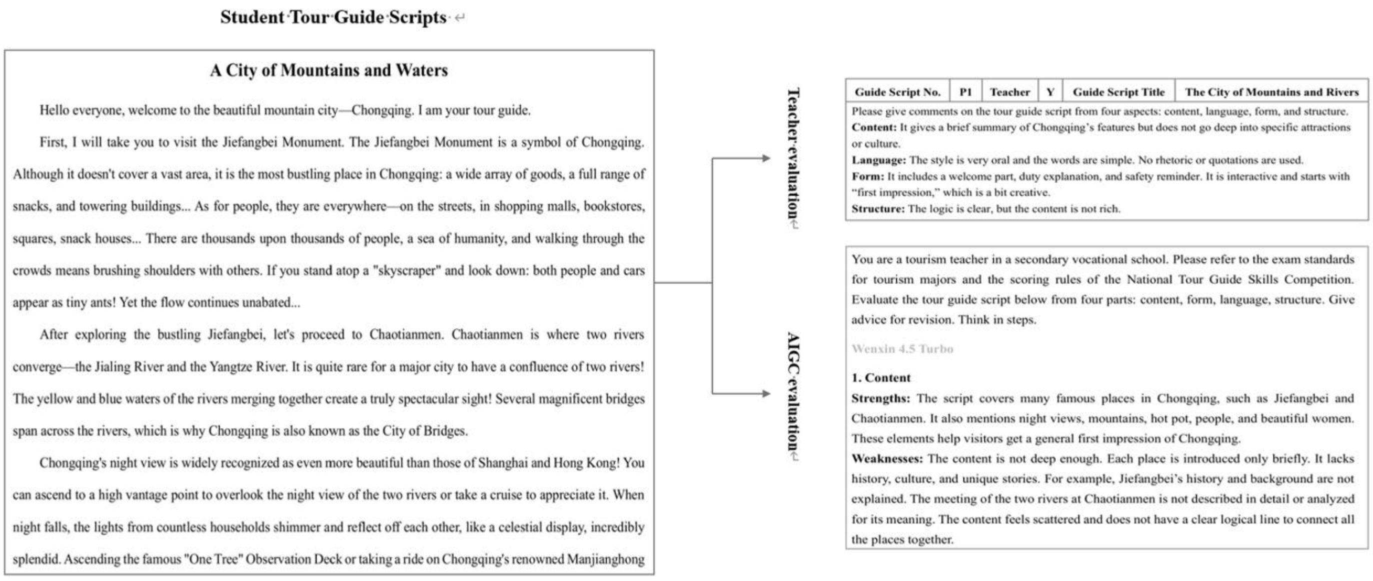
This study employs random sampling to select 152 students from Chongqing Longmenhao Vocational School and Chongqing Tourism School in Chongqing, China, as research subjects. From their assignments, 100 tour-guide scripts are randomly chosen to form the research corpus. Additionally, five in-service tourism teachers from vocational schools in Jiangsu and Sichuan, China, are randomly selected. All of these teachers have experience in exam grading and are familiar with technology-assisted evaluation methods. They serve as human teacher evaluators in the study.

3.2.2. Research Process

This study employs Baidu's "ERNIE Bot" as the AIGC tool to evaluate tour-guide scripts. Based on its official "Precision Questioning" guide, the prompt is structured as follows: "As a tourism instructor at a secondary vocational school, please consult the 'Academic Exam Standards for Tourism Students in Secondary Vocational Schools' along with the 'Scoring Criteria for the National Tour Guide Skills Competition.' Evaluate the above tour-guide script from four dimensions: content, form, language, and structure, and provide revision suggestions." Please think step by step (Tlili et al., 2023). At the same time, the model is provided with the "Academic Exam Standards for Tourism Students in Secondary Vocational Schools" and the "National Tour Guide Skills Competition Scoring Criteria" (Wang, 2019), which specify the detailed requirements for the four dimensions. To ensure feedback consistency, only the model's first response is retained.

In the human evaluation stage, five teachers (Y, E, S, L, W) are invited to independently evaluate 20 tour-guide scripts. The evaluation dimensions are consistent with the model, and no additional guidance is provided. The evaluation format is shown in Figure 3. To reduce individual bias, the five teachers conduct a cross-review after the evaluation. They reach a consensus on any ambiguous items, and the fifth author of this study summarizes and organizes the results.

Then, the evaluation results from “ERNIE Bot” are given to the teachers and compared with their self-assessments. Building on this, semi-structured interviews are carried out to investigate the features and quality of the evaluation information related to the model.



Student Tour Guide Scripts

A City of Mountains and Waters

Hello everyone, welcome to the beautiful mountain city—Chongqing. I am your tour guide.

First, I will take you to visit the Jiefangbei Monument. The Jiefangbei Monument is a symbol of Chongqing. Although it doesn't cover a vast area, it is the most bustling place in Chongqing: a wide array of goods, a full range of snacks, and towering buildings... As for people, they are everywhere—on the streets, in shopping malls, bookstores, squares, snack houses... There are thousands upon thousands of people, a sea of humanity, and walking through the crowds means brushing shoulders with others. If you stand atop a "skyscraper" and look down: both people and cars appear as tiny ants! Yet the flow continues unabated...

After exploring the bustling Jiefangbei, let's proceed to Chaotianmen. Chaotianmen is where two rivers converge—the Jialing River and the Yangtze River. It is quite rare for a major city to have a confluence of two rivers! The yellow and blue waters of the rivers merging together create a truly spectacular sight! Several magnificent bridges span across the rivers, which is why Chongqing is also known as the City of Bridges.

Chongqing's night view is widely recognized as even more beautiful than those of Shanghai and Hong Kong! You can ascend to a high vantage point to overlook the night view of the two rivers or take a cruise to appreciate it. When night falls, the lights from countless households shimmer and reflect off each other, like a celestial display, incredibly splendid. Ascending the famous "One Tree" Observation Deck or taking a ride on Chongqing's renowned Manjianghong

Guide Script No.	P1	Teacher	Y	Guide Script Title	The City of Mountains and Rivers

Please give comments on the tour guide script from four aspects: content, language, form, and structure.
Content: It gives a brief summary of Chongqing’s features but does not go deep into specific attractions or culture.
Language: The style is very oral and the words are simple. No rhetoric or quotations are used.
Form: It includes a welcome part, duty explanation, and safety reminder. It is interactive and starts with “first impression,” which is a bit creative.
Structure: The logic is clear, but the content is not rich.

You are a tourism teacher in a secondary vocational school. Please refer to the exam standards for tourism majors and the scoring rules of the National Tour Guide Skills Competition. Evaluate the tour guide script below from four parts: content, form, language, structure. Give advice for revision. Think in steps.

Wenxin 4.5 Turbo

1. Content
Strengths: The script covers many famous places in Chongqing, such as Jiefangbei and Chaotianmen. It also mentions night views, mountains, hot pot, people, and beautiful women. These elements help visitors get a general first impression of Chongqing.
Weaknesses: The content is not deep enough. Each place is introduced only briefly. It lacks history, culture, and unique stories. For example, Jiefangbei’s history and background are not explained. The meeting of the two rivers at Chaotianmen is not described in detail or analyzed for its meaning. The content feels scattered and does not have a clear logical line to connect all the places together.

Figure 3. Evaluation format of tour-guide scripts by AIGC and teachers.

3.2.3. Data Collection and Analysis

3.2.3.1. Data Collection and Analysis of Evaluation Counts for Each Dimension of Tour-Guide Scripts

To analyze the evaluation scores of AIGC and teachers across the four dimensions of content, language, structure, and form, a comparative study was conducted. The evaluation data from 100 tour-guide scripts (P1 to P100) were broken down into evaluation units, with each unit representing an independent statement addressing a single issue or text feature as the smallest unit. Two researchers independently followed the same procedure to split and compare 10 randomly selected samples to ensure consistency. After resolving any discrepancies through discussion, they applied the same standard to process all texts into units. A total of 1,891 evaluation units were collected, comprising 1,285 from AIGC and 606 from teachers. After removing 318 duplicate units, 1,573 valid evaluation units remained, as referenced by Hayes and Berninger (2009). In the analysis of quantities, the median was used to represent the number of evaluation units per tour-guide script, given that the distribution of unit counts did not follow a normal distribution. The Mann-Whitney U test was employed to examine differences between AIGC and teachers. Additionally, the effect size, r, was calculated to indicate the magnitude of the difference, with values closer to ±1 signifying a larger effect.

Table 1. Coding table for secondary vocational tour-guide script evaluation information.

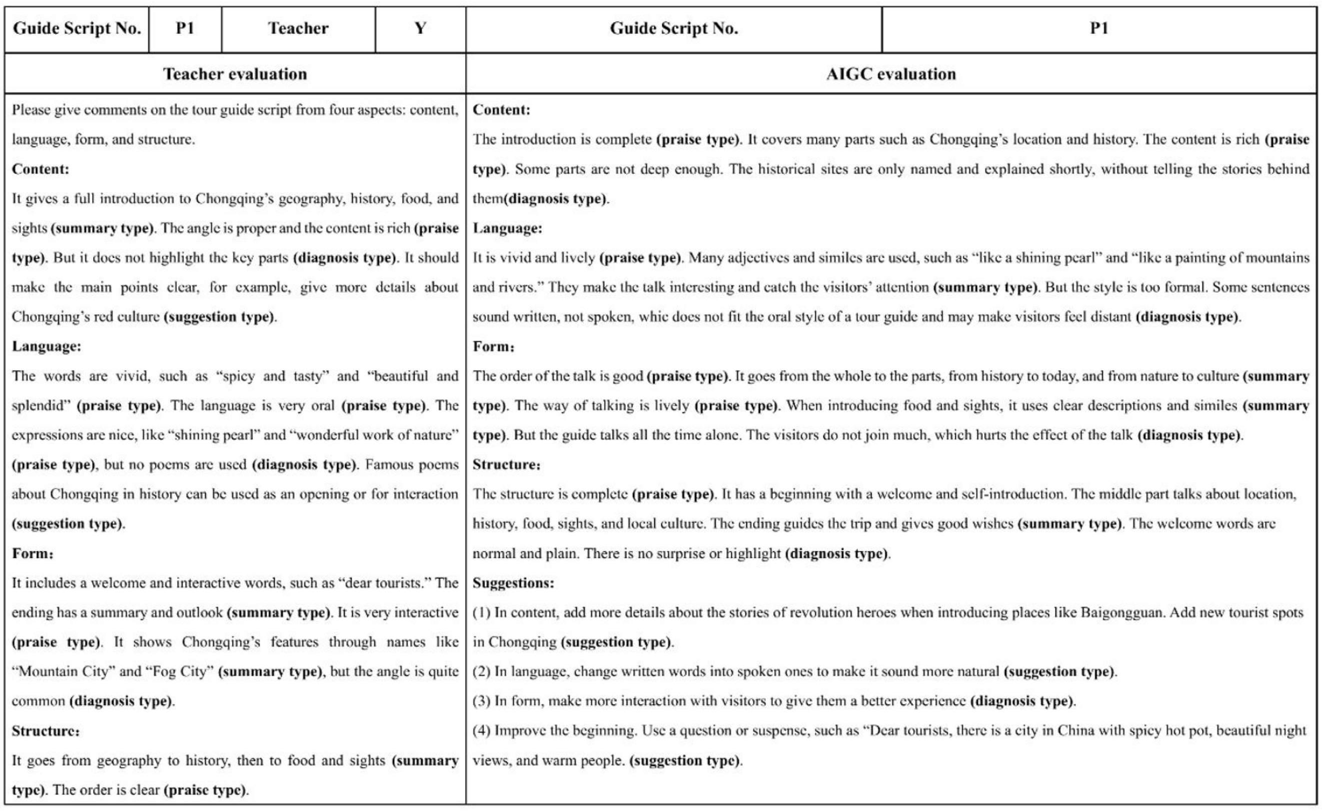
Category	Explanation	Example
Suggestion type evaluation	Point out the mistakes, correct them directly, or tell how to fix them.	“When introducing Chongqing’s features, you can add some history or cultural stories about the People’s Great Hall to make the transition more natural.” “When describing natural sights such as the Huayan Eight Scenes, you can incorporate simple metaphors or personification to enhance the vividness of the description.”
Diagnostic type evaluation	Point out the mistakes, but do not tell how to fix them.	“The introduction of some special sights, such as rare animals and plants, is too general and lacks specific details.” “Some sentences are long and complex, which makes them hard to understand.”
Summary type evaluation	Objectively summarize the writer’s ideas and methods.	“The beginning welcomes visitors and introduces the tour. The main part gives details about Chongqing’s features. The ending shows welcome and good wishes. The logic is clear.” “The beginning welcomes visitors and introduces the tour. The main part gives details about Chongqing’s features. The ending shows welcome and good wishes. The logic is clear.”
Praise type evaluation	Give praise and encouragement for the writing attitude and quality.	“The language is smooth.” “The content is quite complete.” “The logic is clear.” “The structure is clear and well-organized.”

3.2.3.2. Data Collection and Analysis of Evaluation Types

This study first organizes a workshop to systematically review AIGC and teacher evaluations of tour-guide scripts, determining their fundamental categories and features. Next, it adopts the commonly used method for classifying writing assessment information by Wilson and Czik (2016) and after thorough discussion and revision with five teachers, the “Secondary Vocational Tour-Guide Script Evaluation Information Coding Table” is finalized (see Table 1).

In the coding stage, the two researchers who previously handled the splitting task are responsible again. To ensure rule consistency, the two coders first randomly select 50% of the 1,573 valid evaluation units as samples and perform a trial coding based on the “Secondary Vocational Tour-Guide Script Evaluation Information Coding Table” (as shown in Figure 4). After coding, the classification consistency coefficient is used to test reliability. The results show that both coders have a classification consistency index (CA) greater than 0.75, and a reliability coefficient (R) of 0.84 or higher, indicating high coding consistency. This enables a more in-depth analysis.

To compare the evaluation types of AIGC and teachers, the median was used to improve the reliability of the results. Since the proportion data do not follow a normal distribution, the Mann-Whitney U test was employed to assess the differences, and the effect size (r) was provided to indicate the magnitude of the actual differences.



Guide Script No.	P1	Teacher	Y	Guide Script No.	P1
Teacher evaluation				AIGC evaluation	
Please give comments on the tour guide script from four aspects: content, language, form, and structure. **Content:** It gives a full introduction to Chongqing's geography, history, food, and sights **(summary type)**. The angle is proper and the content is rich **(praise type)**. But it does not highlight the key parts **(diagnosis type)**. It should make the main points clear, for example, give more details about Chongqing's red culture **(suggestion type)**. **Language:** The words are vivid, such as "spicy and tasty" and "beautiful and splendid" **(praise type)**. The language is very oral **(praise type)**. The expressions are nice, like "shining pearl" and "wonderful work of nature" **(praise type)**, but no poems are used **(diagnosis type)**. Famous poems about Chongqing in history can be used as an opening or for interaction **(suggestion type)**. **Form:** It includes a welcome and interactive words, such as "dear tourists." The ending has a summary and outlook **(summary type)**. It is very interactive **(praise type)**. It shows Chongqing's features through names like "Mountain City" and "Fog City" **(summary type)**, but the angle is quite common **(diagnosis type)**. **Structure:** It goes from geography to history, then to food and sights **(summary type)**. The order is clear **(praise type)**.				**Content:** The introduction is complete **(praise type)**. It covers many parts such as Chongqing's location and history. The content is rich **(praise type)**. Some parts are not deep enough. The historical sites are only named and explained shortly, without telling the stories behind them**(diagnosis type)**. **Language:** It is vivid and lively **(praise type)**. Many adjectives and similes are used, such as "like a shining pearl" and "like a painting of mountains and rivers." They make the talk interesting and catch the visitors' attention **(summary type)**. But the style is too formal. Some sentences sound written, not spoken, whic does not fit the oral style of a tour guide and may make visitors feel distant **(diagnosis type)**. **Form:** The order of the talk is good **(praise type).** It goes from the whole to the parts, from history to today, and from nature to culture **(summary type)**. The way of talking is lively **(praise type)**. When introducing food and sights, it uses clear descriptions and similes **(summary type)**. But the guide talks all the time alone. The visitors do not join much, which hurts the effect of the talk **(diagnosis type)**. **Structure:** The structure is complete **(praise type)**. It has a beginning with a welcome and self-introduction. The middle part talks about location, history, food, sights, and local culture. The ending guides the trip and gives good wishes **(summary type)**. The welcome words are normal and plain. There is no surprise or highlight **(diagnosis type)**. **Suggestions:** (1) In content, add more details about the stories of revolution heroes when introducing places like Baigongguan. Add new tourist spots in Chongqing **(suggestion type)**. (2) In language, change written words into spoken ones to make it sound more natural **(suggestion type)**. (3) In form, make more interaction with visitors to give them a better experience **(diagnosis type)**. (4) Improve the beginning. Use a question or suspense, such as "Dear tourists, there is a city in China with spicy hot pot, beautiful night views, and warm people. **(suggestion type)**.	

Figure 4. Example of AIGC and Teacher Evaluation Information Splitting and Coding.

3.2.3.3. Data Collection and Analysis of Teacher Perception of AIGC Tour-Guide Script Evaluation Quality

In order to gain a deeper understanding of teachers' opinions on the assessment of AIGC-assisted tour guide scripts, this study conducted semi-structured interviews, each lasting an average of 30 minutes. The interview questions primarily included: "What do you think are the characteristics of the tour-guide script evaluation information from ERNIE Bot?" How do you assess the quality of the tour-guide script evaluation information from ERNIE Bot? What is your opinion on AIGC-assisted evaluation of tour-guide scripts by vocational school teachers? In this study, the interview recordings that were collected are transcribed into text. Thematic analysis was conducted following the steps proposed by Braun and Clarke (2006). Initially, we review the interview content to gain an understanding, and then import the textual data into NVivo 12 for coding and classification. Next, we summarize and name the themes, and finally, we identify the themes in the teachers' responses from the perspectives of strengths and weaknesses.

3.3. Practical Results and Analysis

3.3.1. Comparison of Evaluation Counts by AIGC and Teachers in Each Dimension of Tour-Guide Scripts

Table 2 presents the results of the Mann-Whitney U test comparing AIGC and teachers across the content, language, structure, and form dimensions of tour-guide scripts. The table indicates that AIGC emphasizes structure and form more than content and language. Teachers' evaluations are more evenly distributed across all four dimensions. Additionally, the evaluation counts for AIGC in all four dimensions are significantly higher than those of the teachers. Regarding the magnitude of the differences, the structure dimension shows the largest disparity, followed by form and content, with the language dimension exhibiting the smallest difference.

Table 2. Mann-Whitney U Test Results for Evaluation Dimensions

Evaluation type	Median (range)		U value	Z value	P value	r value
	AIGC	Teacher				
Suggestion	2（2～3）	1（1～1.75）	1753.000	-8.340	P<0.001	-0.589
Diagnostic	2（2～3）	1（0.25～2）	1438.000	-9.815	P<0.001	-0.694
Summary	3（2～3)	1（1～2）	1402.000	-9.081	P<0.001	-0.642
Praise	4（2～5）	1（1～2）	1300.500	-9.344	P<0.001	-0.661

Note: P<0.05.

3.3.2. Comparison of AIGC and Teacher Evaluation Types in Each Dimension of Tour-Guide Scripts

Table 3 presents the results of the Mann-Whitney U test for the evaluation types of AIGC and teachers in the content dimension of tour-guide scripts. As indicated in the table, AIGC predominantly offers suggestion-based and summary-type evaluations, whereas teachers are more inclined to provide diagnostic and praise-oriented evaluations. AIGC's frequency of suggestion-based and summary-type evaluations is notably higher than that of teachers, with a substantial difference in magnitude. Teachers have significantly more diagnostic and praise-based evaluations than AIGC, with a medium actual difference for diagnostic evaluations and a small difference for praise-based evaluations.

Table 3. Mann-Whitney U Test Results for Evaluation Types in the Content Dimension of Tour-Guide Scripts

Evaluation type	Median (range)		U value	Z value	P value	r value
	AIGC	Teacher				
Suggestion	1（0～1）	0（0～0.5）	3584.500	-3.967	P<0.001	-0.280
Diagnostic	0（0～1）	1（0～2）	5983.000	2.604	P=0.009	0.184
Summary	1（1～2）	1（0～1）	3428.000	-4.137	P<0.001	-0.292
Praise	1（0～2）	1（0～3）	5856.00	2.163	P=0.031	0.152

Note: P<0.05.

Table 4 presents the results of the Mann-Whitney U test for evaluation types in the language dimension of tour-guide scripts by AIGC and teachers. As the table illustrates, AIGC generally favors suggestion-based, diagnostic, and summary-oriented evaluations, while teachers are more inclined toward praise-based evaluations. AIGC has significantly more suggestion-based and summary-type evaluations than teachers, with a large actual difference. It also has more diagnostic evaluations than teachers, with a medium actual difference. In contrast, teachers have significantly more praise-based evaluations than AIGC, with a large actual difference.

Table 4. Mann-Whitney U Test Results for Evaluation Types in the Language Dimension of Tour-Guide Scripts

Evaluation type	Median (range)		U value	Z value	P value	r value
	AIGC	Teacher				
Suggestion	1（1～1.75）	0（0～0.75）	2560.000	-6.459	P<0.001	-0.456
Diagnostic	1（0～2）	0（0～1）	3868.500	-3.068	P=0.002	-0.216
Summary	1（1～2）	0（0～1）	2053.000	-7.795	P<0.001	-0.551
Praise	1（0.25～2）	1（0～2）	4108.500	-2.280	P=0.023	-0.161

Note: P<0.05.

Table 5 presents the results of the Mann-Whitney U test for different evaluation types in the Structure Dimension of Tour-Guide Scripts by AIGC and teachers. From the table, AIGC provides more suggestion-type, summary-type, and praise-type evaluations. Teachers, on the other hand, give more diagnosis-type evaluations. In terms of quantity, AIGC offers significantly more suggestion-type, summary-type, and praise-type evaluations than teachers. Among these, the difference in summary-type evaluations is particularly large. The differences in suggestion-type and praise-type evaluations are moderate. Conversely, teachers provide considerably more diagnosis-type evaluations than AIGC, with a large difference observed.

Table 5. Mann-Whitney U Test Results for Evaluation Types in the Structural Dimension of Tour-Guide Scripts

Evaluation type	Median (range)		U value	Z value	P value	r value
	AIGC	Teacher				
Suggestion	1（0～1.75）	0（0～1）	4176.000	-2.184	P=0.029	-0.154
Diagnostic	0（0～3）	0（0～1）	3929.500	-2.817	P=0.005	-0.199
Summary	1（0～2）	0（0～1）	3397.00	-4.228	P<0.001	-0.298
Praise	1（0～1）	0（0～1）	4072.500	-2.517	P=0.012	0.177

Note: P<0.05.

Table 6 displays the results of the Mann-Whitney U test for evaluation types in the form dimension of tour-guide scripts, comparing AIGC and teachers. According to the table, AIGC is more likely to offer suggestion-based, summary-type, and praise-oriented evaluations, while teachers tend to favor diagnostic evaluations. In terms of quantity, AIGC significantly surpasses teachers in these three evaluation types. The actual differences in magnitude are large for summary-type and praise-based evaluations and medium for suggestion-based evaluations. Conversely, teachers have significantly more diagnostic evaluations than AIGC, with a large actual difference.

Table 6. Mann-Whitney U Test Results for Evaluation Types in the Form Dimension of Tour-Guide Scripts

Evaluation type	Median (range)		U value	Z value	P value	r value
	AIGC	Teacher				
Suggestion	1（0～1）	0（0～1）	4072.500	-2.485	P=0.013	-0.175
Diagnostic	1（0～2）	1（0～1）	3852.000	-2.956	P=0.003	-0.209
Summary	1（1～2）	0（0～1）	1802.500	-8.379	P<0.001	-0.592
Praise	1（1～2）	0（0～1）	2233.000	-7.255	P<0.001	0.513

Note: P<0.05.

3.3.3. Teacher Perception of AIGC Tour-Guide Script Evaluation Quality

Through thematic analysis of teacher interview data, the table indicates that in the evaluation of tour-guide scripts, AIGC offers advantages such as clear guidance, high-level strategies, objectivity, comprehensiveness, detailed expression, and accurate targeting. However, it also has some drawbacks, including overly similar expression methods, insufficient alignment with the content and textual context, and poor stability and consistency of the output (see Table 7). All five teachers agreed that AIGC is an effective supplementary method for teacher evaluation and can improve the efficiency of the evaluation process. However, they also emphasized that professional judgment by teachers is necessary during this process to ensure the quality of AIGC's output and to select only accurate and relevant information for the final evaluation.

Table 7. Teacher Perception of AIGC Tour-Guide Script Evaluation Quality

Advantage	Explanation	Example
Clear guidance	The revision suggestions are detailed, specific, and practical.	The suggestions of AIGC are very detailed and specific. For instance, in P1, it states: "Divide long paragraphs into several short ones. Each paragraph should focus on one topic. Add transitional words to make the speech more fluent. Add interactive parts, such as questions or short stories, to increase the visitors' interest and participation." (Teacher S)
Advanced strategy	Provide guidance on advanced writing skills, which is inspiring.	AIGC's suggestions can improve the expression of the theme. For example, in P54, it suggests: "Give a deeper introduction. For the Buddhist story sculptures, choose one or two typical stories and explain the Buddhist ideas and life lessons they show." （Teacher L）
Rational and objective	The evaluation is fair and objective, without personal feelings.	AIGC's evaluation standard is stable and not influenced by personal preference. (Teacher Y)
Comprehensive and detailed	The evaluation is comprehensive and can make up for the details that teachers may miss.	AIGC's evaluation is more comprehensive and detailed. It has both positive aspects and shortcomings for each part, thus making it more rigorous. For instance, it will provide detailed evaluations of both advantages and disadvantages, and also offer relevant suggestions. (Teacher Y, Teacher E)
Accurate evaluation	The evaluation is accurate and can clearly point out the weaknesses.	AIGC can accurately point out problems in logic and grammar. For example, in P31, it suggests: "The word 'first' is repeated. This is a wrong expression and affects the correctness and fluency of the language." （Teacher E）
Weakness	**Explanation**	**Example**
The comments are simple and lack variety	The comments are straightforward and follow a fixed pattern, lacking specificity.	AIGC gives similar comments in many tour guide texts. For example: "The tour guide text is rich in content, with full and detailed information." "The language is accurate and clear, and can express the information correctly." （Teacher S）。
Content mismatch	The evaluation content is not consistent with the local culture.	The evaluation content from AIGC is not consistent with the related content or answers in the textbook. (Teacher E）。
Poor stability	For the same problem, it does not always provide the same judgment and sometimes yields different evaluations.	Both P61 and P72 have some parts that are plain and lack attraction and interest, but AIGC found the problem in P61 and did not find it in P72. （Teacher L）

The research results indicate that AIGC demonstrates significant advantages in the assessment of tour guide scripts. It effectively addresses issues related to insufficient coverage caused by teachers' limited time and lack of concentration, particularly in terms of content richness, language accuracy, structural logic, and innovative presentation formats. With its powerful information processing and pattern recognition capabilities, the artificial intelligence-assisted teaching system (AIGC) can complement teachers' instructional strategies and provide a more comprehensive, detailed, and diverse perspective for assessment. However, AIGC still has limitations in semantic understanding, content extraction, and key information identification. Its sensitivity to context and the accuracy of focus need to be improved.

Based on this, it is recommended that an operational model of "teacher-led, artificial intelligence-assisted" human-machine collaboration be established. Teachers should utilize their professional judgment to screen and correct content generated by artificial intelligence, thereby fully leveraging the strengths of both to improve the scientific accuracy, correctness, and comprehensiveness of the assessment of tour guide scripts.

4. Conclusion

Based on the explanation of the TPACK concept, this study developed the "SICMCS Six-Step Method" for the evaluation and revision of tour guide scripts. The method includes six stages: setting standards, standardized input, AIGC check, guided modification, comparative analysis, and summary sharing. Subsequently, this model was applied in the assessment script practice of the tour guide course. Through content analysis, the differences in the information provided by various assessment sources were compared, while thematic analysis was used to deeply interpret the teacher interview texts. The results indicate that AIGC and teachers exhibit significant complementarity in assessment strategies, which can effectively enhance assessment efficiency. However, there remains room for improvement in the depth and accuracy of the assessments. This research provides a feasible pathway for the in-depth application of technology that combines artificial intelligence and gaming in teaching tour guide scripts, and it also offers relevant empirical evidence.

References

AbuSeileek, A. F. (2013). Using track changes and word processor to provide corrective feedback to learners in writing. *Journal of Computer Assisted Learning, 29*(4), 319-333. https://doi.org/10.1111/jcal.12004

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cao, H. (2022). Research on problems and countermeasures in primary school Chinese composition evaluation and revision. *Language Construction, 2022*(10), 73–74.

Chen, C.-F. E., & Cheng, W.-Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112.

Chi, S., & Shi, J. (2012). Design and application of college physics experiment teaching based on the TPACK framework. *Modern Educational Technology, 22*(8), 33–35.

Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems, 2*(2), 81-83. https://doi.org/10.54097/fcis.v2i2.4465

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97-107. https://doi.org/10.1093/elt/ccn023

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*, 8435-8463. https://doi.org/10.1007/s10639-023-12146-0

Hayes, J. R., & Berninger, V. W. (2009). *Traditions of writing research*. New York: Routledge.

Hoang, G. T. L., & Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of my access. *Language Assessment Quarterly*, *13*(4), 359-376. https://doi.org/10.1080/15434303.2016.1230121

Huang, J. (2019). *Research on the TPACK development of primary school teachers in Wuhan*. Wuhan: Wuhan University.

Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, *32*(2), 131–152. https://doi.org/10.2190/0EW7-01WB-BKHL-QDYV

Li, H. L. M. (2020). A study on the application of peer evaluation in writing in blended learning. *Research on Educational Technology*, *2020*(6), 83-90.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *2015*(27), 1-18.

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

Liu, L. (2019). Research on the transformation of teachers' roles under the TPACK framework. *Theory and Practice of Education*, *19*, 39–42.

Liu, S., Li, Y., & Yang, P. (2021). Research on the effect of intelligent composition evaluation. *Open Education Research*, *3*, 73–84.

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Page, E. B. (1967). *Grading essays by computer: Progress report*. Paper presented at the Proceedings of the Invitational Conference on Testing Problems New York Princeton, NJ: Educational Testing Service.

Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, *70*, 207-221. https://doi.org/10.1016/j.chb.2016.12.076

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Wang, H. (2019). Hierarchical refinement of scoring standards for junior high school entrance examination composition. *Education and Culture Materials*, *33*, 54–55.

Wei, S., & Li, L. (2023). Research on AI-assisted second language writing feedback: A case study of ChatGPT. *Foreign Languages in China*, *20*(3), 33–40.

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94-109. https://doi.org/10.1016/j.compedu.2016.05.004

Zhang, Y., & Wang, X. (2024). Intelligent education application of AIGC: Functional upgrade of intelligent teaching systems. *China Medical Education Technology*, *3*, 312.

Asian Online Journal Publishing Group is not responsible or answerable for any loss, damage or liability, etc. caused in relation to/arising out of the use of the content. Any queries should be directed to the corresponding author of the article.