Journal of Information Technology Education: Innovations in Practice

An Official Publication of the Informing Science Institute
InformingScience.org

JITEiip.org

Volume 25, 2026

LEARNING-ORIENTED ASSESSMENT ENHANCED BY TECHNOLOGY: EFFECTS ON LANGUAGE PROFICIENCY AMONG CHINESE TEVT STUDENTS

Jing Zha *	Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia; & Wuhan Technical University, Wuhan, China	HB220107@student.uthm.edu.my fannyzha@hotmail.com
Azita Binti Ali	Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia	azita@uthm.edu.my
Yan Tan	Wuhan Technical University, Wuhan, China	563749690@qq.com
Yuting Wang	Wuhan Technical University, Wuhan, China	150284134@qq.com

* Corresponding author

ABSTRACT

Aim/Purpose	This mixed-methods study investigates the effect of a technology-enhanced, learning-oriented assessment (LOA) intervention on the language proficiency of Technical and Vocational Education and Training (TVET) students in China. It seeks to uncover LOA strategies that could effectively enhance students' language learning outcomes in flipped classroom instruction, with special focus on the role of technology in mediating LOA practices. By synthesizing quantitative and qualitative data, this study aims to furnish substantial empirical evidence to validate LOA practices in the specific context of language teaching in TEVT and offer valuable insights into the optimization of technology-enhanced assessment practices within the flipped classroom contexts.
Background	Learning-oriented assessment prioritizes learning and demonstrates great potential in enhancing learning outcomes and fostering self-directedness. However, concerns have been raised about the practicality of the LOA and its

Accepting Editor Faisal Syafar | Received: October 10, 2025 | Revised: January 4, January 4, February 12, 2026 | Accepted: February 18, 2026. J
Cite as: Zha, J., Ali, A. B., Tan, Y., & Wang, Y. (2026). Learning-oriented assessment enhanced by technology: effects on language proficiency among Chinese TEVT students. *Journal of Information Technology Education: Innovations in Practice, 25,* Article 14. https://doi.org/10.28945/5727

claimed benefits across various contexts owing to reported implementation challenges and inconsistent evidence of its effectiveness in improving learning outcomes. These challenges can be addressed by improving the design of LOA interventions and enhancing the effectiveness of technological support.

Methodology

A pretest–posttest between-group quasi-experimental design, complemented by focus group discussions and classroom observations, was employed to examine the effects of LOA on language proficiency and to identify effective implementation strategies. The LOA intervention was developed based on Carless's (2007) LOA principles and Jones et al.'s (2016) LOA cycle. Four intact classes were randomly assigned to either the experimental or control group. The modified PRETCO-A test, which had been verified for reliability prior to the intervention, was administered to both groups before and after the 12-week treatment period. Paired-sample and independent-sample t-tests were conducted using the Statistical Package for the Social Sciences (SPSS) to compare the mean scores of overall language proficiency and individual subskills within and between groups. Data from the focus groups and classroom observations were analyzed using content and descriptive analyses.

Contribution

This study contributes empirical evidence on the validity of LOA practice by revealing its positive effects on the specific language skills of TEVT students in an exam-oriented educational setting. By identifying the language skills most strongly enhanced by the LOA intervention and elucidating the challenges encountered and the practical strategies adopted in a flipped classroom, the findings offer valuable insights for researchers and educators seeking to optimize LOA implementation and technology-mediated assessment practices within a flipped classroom context. Particularly, this study advances the understanding of the role of technology in enhancing the LOA process by illustrating how digital tools can integrate assessment, instruction, and learning; enable an action-oriented feedback loop; support more effective assessment task design; and promote rubric-driven peer assessment. Furthermore, it expands the discussion on how the affordance of technology can be leveraged to bolster LOA without compromising academic honesty and fostering student disengagement.

Findings

Although the independent-samples T tests demonstrated no statistically significant overall gains in TEVT students' language learning outcomes following the LOA intervention, a considerable post-test difference emerged in writing skills (writing, $p=0.039$), with a small effect size ($\eta 2=0.0363$), and no corresponding pre-test difference. Qualitative evidence from focus group discussions and classroom observations indicated that effective LOA strategies prioritized rubric-driven peer assessment, an action-oriented feedback loop integrating individualized e-feedback and teacher guidance, and protocol-based self-assessment and self-reflection. While technology has enhanced the effectiveness and efficiency of the LOA procedure, teachers should remain alert to the risks of academic dishonesty and disengagement associated with technological use.

Recommendations for Practitioners

When designing LOA tasks, teachers should align them with TVET students' prior knowledge, desired learning outcomes, and future professional needs, ensuring appropriate challenges and relevance. Prior to peer or self-assessment, students should receive training on how to provide effective feedback and this practice needs to be supported by incentives to foster engagement. Well-structured, rubric-driven peer-assessment and protocol-based self-

assessment and reflection protocols are recommended to reduce cognitive load and enhance participation. Additionally, instant diagnostic feedback and scaffolding should be provided to students to foster a growth mindset and diminish a grade-oriented mindset. Moreover, despite technological affordances, consistent teacher guidance remains crucial throughout the LOA process.

Recommendations for Researchers: The research suggests that the impact of LOA may be limited in scope and may vary across dimensions of language proficiency and among diverse learner types. Researchers are recommended to investigate the impact of prolonged interventions across different learner types and diverse pedagogical settings. Although the current study provides valuable quantitative insights, the discussion remains limited in pedagogical scope. More studies are needed to broaden the understanding of technology-enhanced LOA strategies by incorporating learners' perspectives to provide a more comprehensive view of the mediating effects of students' LOA literacy and prior learning experiences on learning outcomes. Moreover, given the growing concerns over academic integrity and student disengagement in technology-mediated assessment observed in this study, researchers are encouraged to explore the strategies that leverage technology to support LOA while simultaneously addressing these issues.

Impact on Society: The findings provide robust evidence for the validity and practicality of the LOA, particularly the effectiveness of rubric-driven peer assessment, feedback loops, and protocol-based self-assessment and reflection in enhancing language proficiency. This study offers educators valuable insights to optimize LOA implementation, foster pedagogical innovation, and contribute to the broader assessment reform promoted by national educational authorities. It also highlights the potential tensions in technology-mediated assessment, emphasizing the need to balance its benefits with safeguards that ensure academic integrity.

Future Research: Future research should focus on developing systematic strategies for implementing LOA by further testing its validity and effectiveness across different learner groups and in diverse pedagogical contexts. In addition, the integration of emerging technologies in LOA warrants thorough investigation, particularly regarding their pedagogical affordances and associated ethical considerations.

Keywords: learning-oriented assessment, technology-enhanced, language proficiency, technical and vocational education and training (TVET)

INTRODUCTION

Assessment is a fundamental component of instruction and the learning process (González et al., 2018; Zhou, 2023) and is widely recognized as a primary factor influencing learning outcomes (Brown, 2019). With the paradigm shift toward learner-centered pedagogy, assessment practices have progressively shifted from judging outcomes through summative assessment to supporting the learning process via formative assessment, and ultimately to fostering learning itself. This has given rise to learning-oriented assessment (LOA), which prioritizes learning by focusing on assessment tasks, student engagement in assessment, and feedback. Proponents argue that by integrating various assessments strategies, such as assessment of learning, assessment for learning, and assessment as learning, the LOA is expected to inform instructional decisions, boost learning motivation, foster higher-order

cognitive skills, and ultimately promote student achievement. However, despite some empirical evidence of LOA's impact on learning outcomes, there remains a lack of a holistic perspective on LOA practice supported by technology across contexts with varying skill levels (Er & Farhady, 2023). To date, research has yielded limited and inconsistent evidence on the effectiveness of LOA in improving students' English language proficiency in Technical and Vocational Education and Training (TVET) institutions in China, where students' diverse educational backgrounds and English proficiency levels pose additional challenges for language learning and assessment. Moreover, scant empirical research has explored how technology-enhanced LOA operates within this context, particularly in relation to students' diverse backgrounds and learning needs.

Despite the potential benefits of learning-oriented assessment, concerns about the practicality of LOA and its claimed benefits in varied contexts have been raised due to the reported challenges in its implementation (Gao, 2017; R. Yang, 2023). Substantial challenges observed in the specific context of English as a foreign language include students' poor assessing expertise (Jalilzadeh & Coombe, 2023), their resistance to unconventional assessment and their mistrust of peer and self-assessment (Fazel & Ali, 2022), teachers' lack of LOA literacy, time budget, administrative duties, and large class size (Derakhshan & Ghiasvand, 2022; R. Yang, 2023; C. G. Zhao & Qi, 2023). Moreover, implementing LOA is particularly challenging in China's large tertiary classrooms because it is time-consuming and labor-intensive for instructors (Quyen & Khairani, 2017). The specific challenges mostly stem from the capture and use of sufficient data that demonstrate the evidence for and of learning, especially data on unplanned and impromptu assessments (Saville, 2021). Until recently, there was a dearth of research that accounted for the overall picture of "learning" (Banerjee, 2021), but technology now offers viable ways to address these problems. However, resolving this issue is far from easy (Saville, 2021).

The integration of technology has reshaped the educational landscape, notably redefining assessments (Awang, 2021). In this context, technology-enhanced LOA is conceptualized not merely as the digitization of testing but also as the strategic use of digital mechanisms to capture developmental data and to scaffold the learning process. Technologies such as automatic scoring, data mining, learning analytics, and adaptive testing can bring many benefits to LOA practitioners, such as ease of tracking learning progress, time saving, providing instant and personalized feedback (Nikou & Economides, 2018), avoiding academic dishonesty, and supporting not only formative assessment and high-stakes summative assessment (Onasanya & Ajamu, 2024) but also alternative assessment approaches such as e-portfolio assessment, continuous assessment, and dynamic assessment. However, existing studies tend to accentuate technology as a tool for enhancing the efficiency and precision of assessments. Simultaneously, its potential to support the design of LOA tasks and practices remains underemphasized (Greenspon, 2023).

These researchers called for validating the practicality and effectiveness of the LOA enhanced by technology across diverse pedagogical contexts. To this end, the LOA model has been constructed and validated in various educational contexts, demonstrating positive effects on oral language proficiency (e.g., Matyakhan et al., 2024), writing skills (e.g., Er & Farhady, 2023; Hamzelou et al., 2022), and linguistic knowledge, especially pronunciation, lexical, and syntactic knowledge (e.g., Navaie, 2018; X. Yang, 2020). However, the research findings are inconclusive. One study indicated that students' language-learning outcomes did not differ significantly (e.g., Viengsang & Wasanasomsithi, 2022). Moreover, a complete and thorough implementation of the LOA framework was rarely observed in these studies, and there is a notable lack of attention to how technology facilitates LOA in specific settings, such as blended learning, online learning, and flipped classrooms (Susanto, 2025).

To further validate the effectiveness of technology-enhanced LOA on the language proficiency of students in the specific context of the flipped classroom in TEVT institutions in China, where the conventional in-class and out-of-class instruction procedures are reversed with the support of technology. A technology-enhanced LOA intervention was designed and implemented. A quasi-experiment was employed to determine the effectiveness of the LOA on students' language proficiency. A

follow-up focus group discussion was convened to examine LOA practices and evaluate the research findings. This mixed-methods research is expected to contribute to the existing body of empirical studies by providing valuable insights into how LOA practice relates to learning outcomes and how technology can be employed to maximize the benefits of LOA in language pedagogy. The findings will help LOA practitioners optimize their practice with the assistance of technology and provide additional empirical evidence of the effectiveness of LOA in enhancing learning outcomes.

Research Questions

This study aims to validate the effectiveness of technology-enhanced LOA intervention on TEVT students' language proficiency. In particular, it intends to examine which specific skills benefit most from LOA practice and the role technology plays in supporting its implementation. Moreover, it seeks to identify LOA strategies that are both practical and effective for improving the language proficiency of vocational college students. To this end, we set out to address the following five research questions.

RQ1. Is there a notable difference in the English language proficiency levels of vocational college students before and after technology-enhanced LOA treatment?

RQ2. Is there a notable difference in language proficiency levels between students in the intervention and control classes before and after the intervention?

RQ3. What is the effectiveness of technology-enhanced LOA on specific language skills?

RQ4. Which specific LOA strategies are effective in improving students' language proficiency?

RQ5. What role does technology play in implementing learning-oriented assessment (LOA)?

Literature Review

The literature on Learning-Oriented Assessment (LOA) frameworks, interventions, effects on language proficiency, and the use of technology is reviewed in the following section, serving as a theoretical guide for designing the study's intervention, rationalizing the research problems, and framing the discussion of subsequent findings.

Learning-Oriented Assessment Framework

Recently, there has been growing interest in integrating the LOA framework into language curricula (M. Ma, 2023). There are three influential LOA models. Carless (2007) created an LOA model with three components: assessment tasks as learning tasks, student involvement in the assessment process, and feedback as feedforward. The model is primarily intended to address passive learners and the influence of exam-oriented learning (Tsushima, 2022). Hamp-Lyons (2017) added two more key elements to the LOA framework: teacher questioning and scaffolding of performance (Hamp-Lyons, 2017). The second model is Jones et al.'s (2016) LOA cycle for language classes. The LOA cycle focuses on assessment tasks and underscores the utility of the information obtained from them (Viengsang & Wasanasomsithi, 2022). According to the cycle, interaction in the classroom context encompasses five basic elements: task, feedback, goal, scaffolding, and emergence (Jones et al., 2016). When engaging in a task, learners must establish specific objectives. Scaffolding refers to the support provided to students to perform learning tasks, while emergence describes the process of generating language utterances. Tasks are important to language learning in the classroom context, as situational tasks create purpose and meaning for language communication and are related to the elicitation dimension of LOA. Feedback on language performance is at the heart of the learning dimension

(Turner & Purpura, 2016). In addition to these two models, Turner and Purpura (2016) proposed an integrated working framework for LOA. The working framework is based on social constructivism and emphasizes the role of social interaction in language learning. It centers on contextual dimension, elicitation dimension, learning dimension, second-language (L2) proficiency dimension, instructional dimension, interaction dimension, and affective dimension. In 2018, Purpura and Turner revised the model and reconceptualized the seven dimensions (Purpura, 2021). The working framework embodies the multifaceted, complex nature of LOA (Viengsang & Wasanasomsithi, 2022). Its significance lies in providing a holistic view through multidimensional assessment (Tsushima, 2022). According to Purpura and Turner's (2018) LOA working model, the validity of LOA practice should be examined across four interrelated dimensions: the extent to which the assessment tasks can elicit valid evidence of learning, the degree to which affective and social factors mediate students' engagement in assessment, and how assessment results are interpreted and feedback can be delivered to inform instruction and promote learning. These dimensions must be considered in LOA interventions and classroom observations. Concurrently, instructors should be trained to ensure they have adequate L2 pedagogical competence and content and topical knowledge (Turner & Purpura, 2016).

In addition to the three prevailing models, Chinese scholars Zeng et al. (2018) proposed a holistic framework for LOA based on a literature review, focusing primarily on its three components. The model is shown in Figure 1. The model illustrates the relationships among assessment of learning, assessment as learning, and assessment for learning, and how they can promote learning and learners' overall development and fulfill the role of certification in education. It also emphasizes the importance of instructors' and students' understanding and experience with assessment.

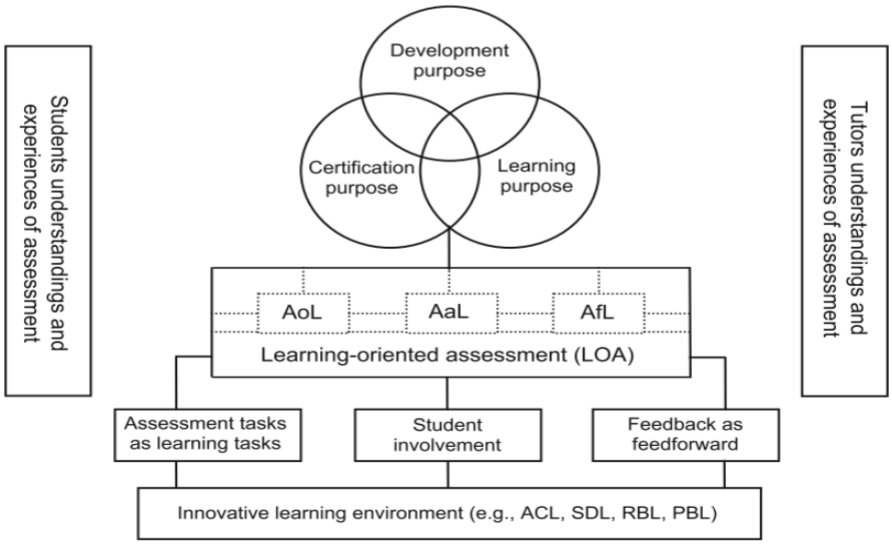


Figure 1. Holistic framework for LOA (Zeng et al., 2018)

LOA frameworks, consisting of assessment for learning, assessment of learning, and assessment as learning, can be used to guide assessment practice in and out of class, facilitating the integration of instruction, learning, and assessment. Despite the well-developed theoretical framework of LOA, more studies are needed in its practical domain (Banitalebi & Ghiasvand, 2023).

LOA Interventions and Their Effects

Theoretically, LOA practice could facilitate learning in several ways. LOA proponents believed that it promoted long-term learning and professional achievement by focusing on tasks and engaging students in assessment and feedback, which helps develop their evaluative skills and, therefore, eases the tension between learning and assessment (Gebril, 2018; Wicking, 2017). Carless (2015) contended

that LOA has positive effects on both the quantity and the quality of interactions (Estaji & Safari, 2023). It helps promote the systematicity of the learning process, making it more curriculum-oriented (Gao, 2017). It can also help learners enhance their autonomy, develop lifelong learning and problem-solving skills, and boost their self-confidence (Bin Mubayrik, 2020; Fazel & Ali, 2022). However, most of these studies employed qualitative or theoretical approaches rather than experimental ones. They focus simply on students' or teachers' perceptions or on researchers' or teachers' hunches. Most of them are not triangulated by multiple sources.

In addition to theoretical arguments regarding the benefits of LOA for general education, empirical studies have demonstrated that LOA can help enhance multilingual language skills (Saville, 2019). Studies have reported that LOA has positive effects on oral proficiency (e.g., Al-Rashidi et al., 2022; Matyakhan et al., 2024), writing skills (e.g., Er & Farhady, 2023; Hamzelou et al., 2022), and linguistic knowledge, especially pronunciation, lexical, and syntactic knowledge (e.g., Navaie, 2018; X. Yang, 2020). Matyakhan et al. (2024) developed a learning-oriented oral communication assessment model that exerted a positive effect on language use, including range, accuracy, fluency, interaction, coherence, and pronunciation. The model highlighted designing learning and assessment tasks and activities based on real-world scenarios, developing students' self-assessment expertise by allowing them to monitor and assess their own performance, and involving students in peer assessment and feedback. The LOA intervention effectively improved Thai students' fluency through self-reflection and peer assessments. The interaction with peers during peer assessment and the repeated giving and receiving feedback from peers were also found to positively impact students' oral fluency (Matyakhan et al., 2024). In a quasi-experimental study conducted by Er and Farhady (2023), self-regulated learning tasks, as well as peer and group tasks, were implemented as primary interventions. The study emphasizes the integration of computer-assisted collaboration, patchwork text assessment, self- and peer-assessment, and feedback mechanisms, which were posited to enhance learners' self-directed learning capacities and enhance their writing performance. Navaie (2018) provided robust empirical evidence that the use of formative assessment for multiple purposes – such as measuring learning outcomes at specific stages, promoting learning, and informing instructional decisions – can effectively enhance language learners' pronunciation skills. The study further highlights the value of engaging both teachers and students in the assessment process and integrating formative and summative approaches, including peer- and self-assessment, to foster sustained improvement in learners' pronunciation performance. Collectively, this body of research underscores the pedagogical potential of formative participatory assessment in fostering learners' metacognitive awareness and improving performance outcomes.

However, the evidence for this is inconsistent. Viengsang and Wasanasomsithi (2022) tested the effectiveness of LOA in improving Thai undergraduate students' reading ability. The in-house language proficiency tests administered before and after LOA implementation exhibited no significant differences. Nevertheless, an analysis of students' journals revealed that the LOA practice may help promote learning by improving learners' confidence, motivation, and self-directedness. According to the research, above better performance in tests, the benefit of the practice lies more in the fact that students are provided with more opportunities to reflect on which language or learning skills need improving and how to improve, what they have learnt, and how far they are from the objectives, which may benefit their lifelong learning (Viengsang & Wasanasomsithi, 2022). Moreover, other studies have claimed to help develop evaluative expertise (Estaji & Safari, 2023; Kim, 2022; Viengsang & Wasanasomsithi, 2022), cognitive knowledge, and metacognitive ability in writing (X. Yang, 2020), as well as skills such as communication and collaboration by increasing classroom interactions, students' engagement, and ability to identify their learning gaps (Viengsang & Wasanasomsithi, 2022).

Use of Technology in LOA

Educational technology is advantageous for delivering learning materials, promoting classroom interaction, and assessing learners (Meccawy et al., 2023). Recent developments in information and

communication technology (ICT) have made the adoption of technology in educational settings more widespread, not only in the learning process but also in assessment (Höl & Akman, 2023). According to Voss (2021), technology can be strategically harnessed to operationalize the LOA's pillar principles in a language learning context.

Specifically, it enhances the efficiency of assessment task development, empowers student involvement in evaluative processes, diversifies the delivery of feedback, and deepens student engagement with it (Voss, 2021). With respect to the principle of designing language assessment tasks for language learning, corpus-based linguistic analysis tools are leveraged to assist instructors in identifying the distinct patterns of the target language and language learners' interlanguage profiles, which not only facilitates instructors in establishing attainable learning objectives but also strengthens the validity of assessment tools in measuring students' ability to use language in authentic contexts (Voss, 2021; Yoqubovna, 2025). For instance, eye-tracking technology provides a window into students' cognitive processes, enabling instructors to refine assessment design by observing learning patterns. Moreover, authoring tools and Learning Management Systems (LMS) optimize the delivery of assessment tasks and the embodiment of inputs through multimedia protocols (Voss, 2021). Regarding student engagement, technology facilitates collaborative peer- and self-assessment by providing multi-user platforms equipped with evaluative scaffolding such as vocabulary profiles, performance rubrics, and automated feedback reports. As for the LOA feedback principle, integrating technology improves the efficiency of delivering feedback and tracking students' responses to it (Voss, 2021).

The most conspicuous affordance of technology in learning-oriented language assessment is its capacity to assist instructors in documenting and monitoring collaborative interaction, thereby enabling the tracking of students' learning progress and their second-language development (Jones et al., 2016). In this way, instructors can better understand how students engage with feedback, allowing them to monitor students' responses and adjust their instructional strategies promptly. However, technologies such as educational data mining and learning analytics have not been fully employed in practice to inform teaching and enhance learning (Guo et al., 2021). Concurrently, there is limited discussion at the pedagogical and governance levels (Butler-Henderson & Crawford, 2020). Further research is needed to determine whether technology-enhanced language assessment effectively facilitates the development of the underlying constructs and fosters better practice in terms of accuracy and authenticity (Sadeghi, 2022).

In China, the incorporation of cutting-edge technologies has significantly transformed the landscape of educational instruction and assessment. To explore the feasibility of the technology-enhanced LOA framework, Chinese researchers conducted experiments in Chinese, mathematics, and English classes taught by artificial intelligence (AI) teachers in a primary school (Ruan et al., 2021). In this research, AI technology served crucial dual functions. First, it is used for comprehensive data collection, gathering quantitative metrics from quizzes and classroom behavioral observations, and analyzing qualitative data derived from system logs and student self-reports. Second, AI is engineered to provide productive, instantaneous feedback. In the classroom, the AI system delivered precise evaluations of student responses, offering instant feedback to both students and educators. This immediate analysis empowers teachers to deliver just-in-time instruction and promptly adjust their pedagogical strategies (Ruan et al., 2021).

The recent emergence of AI models, such as DeepSeek, has enhanced the accessibility of assessment technologies and expanded the potential for LOA. According to Wu (2025), AI is leveraged to generate three-dimensional error analysis reports that diagnose students' misconceptions and map them to specific knowledge gaps for targeted pedagogical remediation in China's primary schools. Based on these assessment results, personalized learning pathways are designed to guide each student's progress toward subject-matter mastery (Wu, 2025). The applications of AI in assessment effectively support the design of assessment tasks, the analysis of learning data, the tracking of learning processes, and the provision of instant feedback (Wu, 2025). However, this focus on automation often neglects crucial pedagogical elements, such as active student involvement in assessment and their

meaningful action upon feedback, which risks fostering superficial learning and a grade-oriented mindset. Thus, it is essential to explore the use of technology to augment the potential of assessment while overcoming constraints to achieve "revolutionary" progress (Chapelle & Douglas, 2006; Voss, 2021).

While technology supports LOA, its implementation is not without challenges. The rise of AI technology invites scrutiny regarding academic integrity and the potential for students to manipulate automated systems for unfair advantages. Studies demonstrate a rise in cheating rate in technology-supported assessment contexts (Janke et al., 2021; Ozdem-Mertens & Orhan, 2023). Students may demonstrate academic dishonesty and a lack of motivation and engagement during formative assessment (Ko & Rossen, 2017). Moreover, Voss (2021) warned that the use of automated writing systems may encourage students to mimic algorithmically favored patterns rather than fostering communicative skills in authentic contexts. Technology integration challenges may include insufficient ICT infrastructure, user anxiety and resistance to change, lack of management support, and loss of personal contact. In this domain, the Unified Theory of Acceptance and Use of Technology (UTAUT), which synthesized eight prominent models into one single framework, can be employed to diagnose and overcome the technology barriers through its four core determinants: performance expectancy, effort expectancy, social influence, and facilitating conditions (Pan & He, 2024). Overall, the application of cutting-edge technology in assessment remains a novel and under-researched area, despite some interest from educators.

Language Proficiency and Pedagogical Practice in China

From the perspective of second-language or foreign-language learning, English language proficiency is a multidimensional construct comprising multiple interrelated components (Yoon & Bhat, 2018). It can be conceptualized as the competence to use the English language to create and convey meaning in both written and spoken contexts (Council of Europe, 2020). Two major trends can be identified in conceptualizations of language proficiency: one foregrounds the underlying and interrelated linguistic and cognitive traits of the language user, whereas the other emphasizes the functional ability to use language effectively to complete tasks in specific contexts (Zúñiga, 2024). Accordingly, these two perspectives are manifested in language proficiency assessment practices, with one approach prioritizing the measurement of learners' linguistic competence and the other focusing on their performance in communicative contexts. While learners' underlying competence tends to be relatively stable, their performance, understood as the manifestation of this latent competence, is often context-specific; therefore, it is reasonable to assess both linguistic competence and communicative skills. In China's vocational colleges, students are encouraged to take the Practical English Test for Colleges (PRETCO), a nationwide standardized English proficiency test created in line with the *Basic Requirements for English Curriculum of Higher Vocational and Technical Education* (Ministry of Education of the People's Republic of China, 2021) and the economic and social needs for language use. It primarily assesses grammatical and syntactic competence as well as listening, reading, writing, and translation skills (Chen, 2022). Unfortunately, despite the pedagogical and administrative endeavors in China's vocational colleges, in general, students' English language proficiency has remained relatively low, manifested by poor communicative skills in oral and written English (J. Li & Paulino, 2025; R. Yang, 2023), a significant disparity among language skills and deficiency in language development (Jiang & Qu, 2024).

To enhance vocational college students' language proficiency and prepare them for future careers, China's *Guidelines on the Formulation and Implementation of Vocational and Technical Training Programs* recommend that colleges adopt innovative instructional approaches, such as flipped classrooms, blended learning, and integrated learning, and integrate technology and digital resources to transform classroom teaching (Yuan & Moran, 2018). In response, technology-assisted flipped classroom models have been operationalized in China's vocational colleges, especially in English classes (Zuo, 2016). Recent research rationalizes the integration of the work process into the curriculum of English for

Special Purpose to improve the TEVT students' employability (J. Ma, 2025), while another study experiments on a project-based flipped learning model to improve TEVT's ability to solve practical problems through effective English communication (Y. Yang & Mohd Ahyan, 2025). These efforts reflect a transformation from traditional, subject-heavy teaching to a model focused on student needs and career readiness. However, teachers also faced significant challenges in localizing the practice. One conspicuous challenge lies in the grade-prioritized mindset and the overemphasis on summative assessment, which are incapable of reflecting students' achievement in skills such as critical thinking, problem-solving, and communication (Yuan & Moran, 2018). While recent research emphasizes curriculum and instructional reform, it often overlooks a critical gap. Despite the shift toward student-centered learning, assessment remains largely teacher-dominated rather than learning-oriented.

METHOD

A proliferation of studies has examined how assessment can facilitate teaching and learning in the language classroom. However, due to its interdisciplinarity, complexity, and rich contextualization, diverse research methods are harnessed to conduct language assessment and LOA research (Tsushima, 2022). As the research aims to explain the cause-and-effect relationship between LOA and language learning outcomes, while also exploring the specific LOA strategies that contribute to these effects, a quasi-experimental pretest–posttest design with non-equivalent groups, complemented by follow-up focus group discussions, was adopted (Creswell & Plano Clark, 2018). In the quantitative phase, a pretest-posttest comparison was conducted to identify potential differences in proficiency test results following the introduction of the LOA invention. As participants generated multiple data points, the means of the two test scores across participants were calculated to represent the intervention effect. However, the pre-test and post-test experimental design could not rule out the possibility that factors other than the intervention may have led to changes in mean scores. Thus, in this research, a between-groups experiment was proposed to address RQ1, RQ2, and RQ3. The qualitative phase was designed to complement the quantitative investigation.

The results of the quasi-experiment were used to inform the plan for the qualitative phase, identifying which questions needed further exploration and who should be involved to provide a good explanation for the quantitative findings (Creswell & Plano Clark, 2018). In this research, the quasi-experiment yielded findings on changes in TEVT students' language proficiency; however, it remains unknown what factors might contribute to these changes. To address RQ4 and RQ5, a focus group discussion was held in which both teachers and observers shared their perspectives on LOA practices. This is because, compared to individual interviews, focus groups allow for more efficient data collection, yielding richer insights (Chang & Hsu, 2006). Teacher participants were asked about their experience implementing the intervention, focusing on which specific LOA strategies were effective in improving students' language proficiency, the challenges they encountered, and the role technology played in LOA practice. Observers' reports and LMS logs were employed to verify the findings from the focus group. Finally, the results from both the quantitative and qualitative phases were discussed, and an integrated conclusion was drawn.

This explanatory sequential mixed-methods design could yield sufficient evidence to delve into and document emerging innovation in assessment practice in authentic pedagogical contexts. Rather than seeking casual generalization through a fully randomized experiment, this research was deliberately framed as an exploratory investigation. It generated inchoate empirical insights into the implementation and pedagogical potential of incorporating LOA into flipped instruction in understudied vocational English classes and in emerging instructional practices. The multiple data sources from the mixed-methods research were likely to yield credible, contextually rich evidence of the workability and instructional value of the innovative practice, as well as the substantial benefits to the students involved. This empirical evidence is adequate to inform LOA practice, sharpen theoretical assumptions, and serve as a source of reference for large-scale or randomized experimental research.

RESEARCH DESIGN

Quantitative phase

In the quantitative phase, the two groups, selected via purposive sampling as outlined in the previous section, are assigned to different treatment conditions. One was taught with the current flipped classroom instructional model, and the other with the LOA-integrated flipped classroom instructional model. Given the most rigorous between-group design, a randomized experiment may disrupt class sessions (Creswell & Guetterman, 2019). Because only intact groups were available to the researcher, a quasi-experimental procedure was implemented. However, to mitigate potential bias inherent in quasi-experimental research, the two groups were randomly assigned to intervention conditions. As shown in Figure 2, the design includes both pre-tests and post-tests, as well as between-groups observations. The intervention conditions were assigned randomly to the four intact classes.

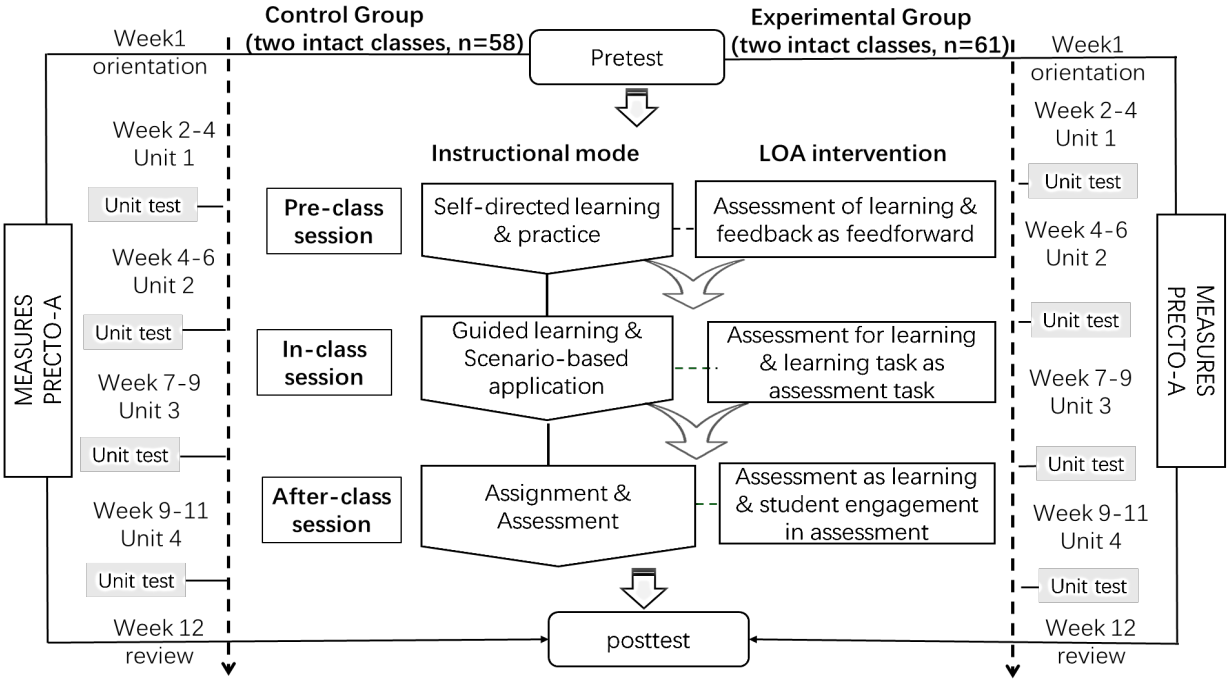


Figure 2. Pretest-posttest between-group quasi-experiment design

To avoid disrupting the normal pedagogical schedule, according to the current semester's academic calendar, the experiment was conducted from week 6 to week 17, when English classes were scheduled. In week 6, orientation was provided to experimental groups to inform them of what the course intended to achieve, what they needed to do to achieve the goal, what resources or software they could access to help them achieve the goal, and how they were to be assessed for their success at the end of the course. The modified PRETCO-A test was administered to both the experimental and control groups before and after the LOA intervention. Data was collected for analysis.

As illustrated in Figure 2, both groups operated under a flipped classroom framework structured into three distinct sessions: self-directed learning and initial practice in a pre-class session; guided learning and scenario-based applications in an in-class session; and assignments and formal assessments in an after-class session. The primary pedagogical variable distinguishing the two groups was the methodology and integration of assessment. In the experimental group, a technology-enhanced LOA framework was practiced. Conversely, the control group used a conventional assessment model comprising standard formative and summative evaluations. As illustrated in Table 1, the experimental and control groups differed in the use of the medium, assessment strategies, assessment tasks and items, and feedback delivery and quality. E-assessment tools such as automatic grading systems, authoring tools,

test editing, settings, e-proctors, and result reporting were used in experimental classes to support bite-sized assessments like vocabulary and grammar quizzes and reading skills tests, provide instant quantitative and qualitative feedback for writing and translation assignments, and record and present assessment results to help instructors and students make informed decisions. Concurrently, AI-generated learner profiles were created to systematically capture students' performance across individual assignments and collaborative activities, including online discussions, group presentations, and classroom interactions, thereby enabling the monitoring of learning progress and identification of learning gaps. In essence, the experimental groups differed from the control groups in that every facet of the assessment served the primary purposes of fostering learning and driving future behavioral change (Gebril, 2021).

Table 1. Comparison of technology and assessment protocol for experimental and control groups

Feature		Experimental group	Control group
Medium	Digital LMS & MOOC	✓	✓
	Interactive tools	✓	✓
	E-assessment tools	✓	×
	AI-generated learner's profile	✓	×
Assessment strategies	Peer assessment	✓	×
	Self-assessment	✓	×
	Product-oriented assessment	✓	✓
	Process-oriented assessment	Used for learning purposes; learning-oriented	To measure students' learning progress, instruction-oriented
	Low-stakes, bite-sized assessment	5-10 minutes prior to every face-to-face session & once for each unit	Once for each unit
Assessment tasks and items	Difficulty	Appropriate and manageable, aligned with feedforward	Consistent to ensure comparability
	Content and type	Varied, interactive, collaborative, and matched with the intended learning outcome	Standardized, primarily written, conventional tests, consistent to ensure comparability
Feedback	Timing	Timely, regularly	Delayed or skipped
	Quality	Actionable, written, qualitative	Oral, quantitative grades
	Reach	Both individual and general via LMS or in class	Only general in class
	Follow-up	Action upon feedback is required and tracked	Action upon it is not necessarily required

Prior to the intervention, teacher participants received training in the core principles and strategies of LOA, with particular emphasis on designing effective assessment tasks and providing constructive feedback within a vocational English curriculum. To isolate the intervention's impact and minimize the influence of extraneous variables, both the experimental and control groups were taught by the same instructors. A shared instructional model, detailed lesson plans, and collaboratively developed assessment rubrics were provided to ensure uniformity in instructional practice and LOA strategies. Furthermore, the two cohorts followed identical curricula, adhered to the same syllabus, used unified instructional materials and assessment tasks, and shared nearly identical timelines across experimental classes. This design ensured that any observed differences in learning outcomes could be explicitly attributed to the nature of the assessment strategies rather than variations in teaching style or core content. Throughout the intervention, regular coordination meetings were held on Thursday afternoons at intervals of three to four weeks to discuss instructional practices and address implementation issues, while periodic classroom observations were conducted by two associate professors and experts in language assessment in a vocational college to monitor conformity to the instructional design and to understand students' responses to the intervention better.

Qualitative phase

Following the quasi-experiment, qualitative exploration was undertaken. In a focus group, two teacher participants shared their experiences with the LOA intervention, including how they implement it in and out of the classroom, their perceptions of students' reactions to the intervention, and their insights into the practice's success and failure. Likewise, the two classroom observers shared their perceptions of teachers' LOA practice, students' responses to teachers' instruction, and reflections on the causes of the experiment's success or failure. The focus group discussion was conducted online via Tencent Meeting, which automatically transcribes and summarizes the key points and each speaker's contributions.

PARTICIPANTS

As the study intended to examine the effect of LOA in the specific context of TEVT English classes in Central China, the study's population was constrained to first-year non-English majors from primarily public TEVT institutions at the tertiary level, which constitute the majority in the region. They were enrolled in three-year vocational and technical programs, majoring in 19 fields of TEVT granted by the Ministry of Education of the People's Republic of China. This cohort of students typically entered vocational or technical colleges with elementary to lower-intermediate English proficiency (approximately CEFR A2-B1), even though they had started learning English in Grade 3 and had been learning it continuously for 6 to 9 years. They gained admission to TEVT institutions primarily through one of the two tracks: the national entrance exam for academic colleges and universities (Gaokao) or the entrance exam for vocational colleges (TEVT Gaokao). They had been exposed to an exam-oriented culture with a strong emphasis on summative assessment for years. At China's TEVT institutions, the assessment and instruction practice of English course are guided by the *English Curriculum Standard for Higher Vocational Education* (Ministry of Education of the People's Republic of China, 2021) issued by the specialized administrative authority. According to the *Curriculum Standard*, the English course is a compulsory course for all first-year students. English classes typically range from 30 to 100 students, and weekly hours are usually limited to 1.5-4.5 hours. However, free online English learning courses are accessible to all students for independent study. The blended instruction model, particularly the flipped classroom model, has been prevalent in TEVT since 2014.

Among the 749 vocational colleges in central China, one vocational college with approximately 4,000 freshmen was selected as the research site. Each class consisted of 30-40 desired participants, which is the average size of vocational colleges and classes in central China. To justify the sample size, the G*power calculator was used to determine the required sample size for the two-group independent-samples t-test. The results showed that 30 participants per group were sufficient for the study. As the intact classes available for the research consisted of around 30 students, two intact classes per group

were an optimal setting for the research, allowing for possible dropouts during the experiment or invalid responses. Students' high school academic backgrounds and English scores in the Gaokao were used as important criteria for selecting potential participants. To avoid extraneous variables, the student participants from the four intact classes were taught by two teachers at two separate campuses of the same vocational college. Consent was obtained from student participants before the experiment. They were well informed about the research objectives, potential benefits, and risks of participating, and their freedom to withdraw at any time. They were assured of the confidentiality of their privacy and that their data would be used only for this research purpose. The four intact classes were randomly assigned to the control group and the experimental group and randomly assigned to be under the instruction of the two teachers by flipping a coin. The details of the student participants are as in Table 2.

Table 2. Detailed information about participants

Class	Group	Faculty	Average age	Gender	Duration of the intervention period	Total credit hours
Pharmaceutical Marketing Class 1	Control group	Biological Engineering	18+	M: 4 F: 26 Total: 30	Null	60
Finance Class 1	Control group	Business	18+	M: 3 F: 25 Total: 28	Null	60
Pharmaceutical Marketing Class 2	Experimental group	Biological Engineering	18+	M: 3 F: 28 Total: 31	12 weeks	60
Finance Class 2	Experimental group	Business	18+	M: 5 F: 25 Total: 30	10 weeks	60

As shown in Table 2, all the student participants aged above 18 years old. Although there was an apparent gender imbalance within the classes, no significant difference in the gender distribution was observed between the experimental groups and the control groups.

The participants of a quasi-experiment involve both the teacher and students. Two teachers from the same vocational college volunteered to participate in the research, both of whom were senior lecturers with over 15 years of experience in English teaching in vocational colleges. Teacher A was more experienced than Teacher B, as teaching had been her primary role since 2007, while Teacher B had other responsibilities in addition to teaching. Moreover, teacher A had been engaged in the pedagogical reform based on the flipped classroom since 2014, while teacher B had less experience with the flipped classroom practice.

INSTRUMENTS

The study selected the modified PRETCO (Level A) as a measure of participants' language proficiency, primarily because the national criterion-referenced English proficiency test is explicitly designed for non-English majors from vocational colleges and aligns with the English curricula of most vocational colleges. In designing the test, the Testing Committee of PRETCO has referred to the *Basic Requirements for English Curriculum of Higher Vocational and Technical Education* (Ministry of Education of the People's Republic of China, 2021), according to which English learners in vocational

colleges should have good English communicative skills to ensure they can meet the challenges in the workplace. PRETCO has been in operation for more than two decades, with approximately 3 million test takers each year (Chen, 2022), and is the most influential and authoritative English language proficiency test among vocational colleges. It is held twice a year, in June and December, respectively, and is divided into two levels: Level A and Level B. The Level A test is the standard requirement for workplace communication, while the Level B test is the minimum requirement (Chen, 2022).

In general, the test is suitable in terms of difficulty and content as a measure of English language proficiency for vocational college students. PRETCO-A consists of five parts, namely listening comprehension, vocabulary and grammar, reading comprehension, translation, and writing. Without modifying the test framework, the researcher subtracts the number of questions in each part to ensure students can finish the test within 1 hour. The modified test includes 39 items. Part 1, reading comprehension, includes 3 short dialogue questions, 2 long dialogue questions, and 10 blanks to fill in a single passage, accounting for 20% of the total score. Part 2, vocabulary and grammar, consists of 10 multiple-choice questions and 5 blank fillings, accounting for 30% of the total score. Part 3, reading comprehension, includes one passage with 5 multiple-choice questions, followed by one piece of practical text with 5 questions, for a total of 25%. Part 4, translation, includes three sentences followed by three choices, for a total of 9%. Part 5, writing, requires students to write a letter, which accounts for 16%. In assessing writing, the researcher refers to IELTS writing rubrics (band description <public version>), in which task achievement accounts for 25%, coherence and cohesion for 25%, lexical resource for 25%, and grammatical range and accuracy for 25%.

To assess the reliability of the language proficiency test, two tests were administered to the same cohort of students at an interval of about 4 weeks; 240 volunteers participated in the tests. The tests were taken online via the LMS; raw data on total scores and scores for each sub-section of listening comprehension, vocabulary and grammar, reading comprehension, and writing were collected and entered into SPSS for analysis. Test-retest reliability procedures were conducted using SPSS to calculate the correlation between the two test scores. The analysis yielded a Cronbach's alpha of 0.861, exceeding the conventional threshold of 0.70. This indicates a high level of internal consistency between the test (pre-test) and retest administration (post-test). The Intraclass Correlation Coefficient (ICC) for single measures was 0.748, with a 95% confidence interval ranging from 0.673 to 0.808, while that for average measures was 0.856, with a 95% confidence interval ranging from 0.804 to 0.894. The F-test results ($F = 6.946$, $p < 0.001$) indicate that these correlations are statistically significant. The results confirm that students' ranking remained consistent throughout the two tests. Thus, it can be concluded that the instrument is a reliable tool and could yield reproducible results with a low margin of measurement error.

STATISTICAL ANALYSIS

In this research, the pretest-posttest between-group quasi-experimental design was used to provide robust evidence on the effectiveness of LOA in improving the language proficiency of TEVT students. To this end, the LOA intervention was administered to two intact classes from different faculties, while two additional intact classes served as the control condition. A preliminary analysis was conducted to verify the assumptions for parametric testing, including tests of normality (Shapiro–Wilk) and homogeneity of variances (Levene's test), using SPSS 26. Upon confirming the suitability of the data, both overall language proficiency scores and sub-skills test scores were analyzed by comparing pre-test and post-test means and the mean differences between the experimental and control groups.

When the data are normally distributed, paired-samples t-tests should be run on data from both the experimental and control groups to determine whether there are differences between pre-test and post-test scores within each group. In the meantime, an independent-samples T-test should be conducted to determine whether there are differences in post-test results between the experimental and control groups, given that the pre-test scores are not significantly different between the two groups.

When data are not normally distributed or are ordinal, nonparametric tests should be used. If the case is valid, the Mann-Whitney test shall be used to assess differences between the control and experimental groups. At the same time, the Wilcoxon Signed Rank test should be used to assess differences between pre-test and post-test scores within each group. When analyzing data, several factors beyond statistical significance should be considered. The effect size should be calculated to evaluate the strength and importance of the difference between the two groups. As t-tests in SPSS do not provide the value of the effect size, G*Power was used to compute effect sizes. Meanwhile, p-value, Type I error, Type 2 error, and power of a test should also be taken into consideration in analyzing and interpreting results.

Following the quasi-experimental phase, a focus group discussion was conducted with teacher participants and classroom observers to obtain qualitative insights into their experiences and perceptions. The discussion was recorded and transcribed using Tencent Meeting, which also generated an automated summary of the session. The resulting qualitative data, including the verbatim transcript and system-generated summary, were systematically analyzed through descriptive and content analysis to identify effective LOA strategies, challenges, and the role of technology, aligned with the study's research objectives.

RESULT

Overall, the mixed study indicates that the LOA intervention had a positive effect on TEVT students' language proficiency; however, the effect did not reach statistical significance. Nevertheless, significant effects were observed on English writing skills, with the experimental groups outperforming the control groups. The follow-up discussion highlighted the positive effect of meticulously designed peer assessment, effective feedback, and well-structured self-assessment and self-reflection in improving students' learning outcomes and self-directed learning skills.

OVERALL EFFECT OF INTERVENTION ON LANGUAGE PROFICIENCY

In this study, the integration of LOA in flipped classroom are assumed to be factors that may explain variation in language proficiency. The LOA embedded instructional mode is the independent variable, and language proficiency level is the dependent variable. The study compared the mean of two groups: the experimental group (with LOA elements) and the control group (without LOA elements) in terms of language proficiency. To do so, first of all, the normality and homogeneity of data collected from both language proficiency tests need to be assessed to determine an appropriate statistical test for further processing.

Preliminary analysis

Prior to performing formal data analysis, several assumptions need to be checked to determine whether the data is suitable for parametric analysis or non-parametric analysis. Language proficiency test scores were measured on a continuous scale with fixed increments between values, but without a true zero point (Powers et al., 2017), which met the first assumption. They were then checked for normality and whether their distribution across both the experimental and control groups is comparable or not (Greasley, 2008). To check normality, in this research, visual inspection was performed using a histogram, and statistical testing was conducted using the Kolmogorov-Smirnov (K-S) or Shapiro-Wilk (S-W) tests (Roni et al., 2019). The normality test results for the students' language proficiency test scores in the study are presented in Table 3.

Table 3. Tests of normality

	Class	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Pretest	Control group	0.110	58	0.176	0.973	58	0.216
	Experimental group	0.079	61	0.200*	0.953	61	0.020
Posttest	Control group	0.097	58	0.200*	0.964	58	0.080
	Experimental group	0.099	61	0.200*	0.970	61	0.133

* This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As shown in Table 3, the significance level for S-W and K-S tests for the two classes was 0.176, 0.2, 0.2, 0.2, and 0.216, 0.020, 0.080, and 0.133, all of which were above 0.05 except for the pretest results of the experimental group. The non-significant value of above 0.05, or less stringently above 0.01, indicated the normal distribution of pretest and posttest scores for both experimental groups and control groups. The results were further supported by the histogram, which showed a slightly skewed distribution, and by the Normal Q–Q plots of the two test scores of the two groups. Most points closely followed the reference line, with only minor deviations at both ends of the pretest distribution. These deviations are acceptable, indicating an approximately normal distribution, particularly given the robustness of the statistical tests employed (S. Yang & Berdine, 2021).

As to the second assumption concerning the homogeneity of variances, SPSS was used for statistical analysis, which provided a built-in correction to address the potential violation of this assumption (Gliner et al., 2017). Thus, in this research, parametric data analysis was run to test the hypothesis.

Paired sample T-test result

Following the Normality Test, inferential statistics were run. The research sets out to test the effectiveness of the LOA intervention on vocational college students' language proficiency. Two versions of PRETCO-A, a high-stakes English proficiency test specifically designed for vocational colleges, were administered to both the control and experimental groups before and after the intervention. To test whether language proficiency improved before and after the intervention, paired-samples t-tests were conducted to compare pretest and posttest mean differences. Prior to the comparison tests, it is essential to ensure that the pretests and posttests were from the same sample. To this end, the correlation tests were conducted. The output generated from the test is as follows in Table 4.

Table 4. Paired samples statistics

		N	Correlation	Sig.
Control group	pretest & posttest	58	0.348	0.007
Experimental group	pretest & posttest	61	0.452	0.000

As shown in Table 4, the P values of the pretest and posttest scores for both the control group and the experimental group were below 0.05, indicating that the pretest and posttest scores were significantly related and came from the same subjects.

Table 5 shows that the mean difference between the pretest and posttest scores was 19.379 for the control group and 21.115 for the experimental group. The two-tailed significance values for both the control group and the experimental group were 0.000 ($p < 0.05$), indicating statistically significant differences between the pretest and posttest scores in both groups. It further indicated that both flipped classroom instruction plus conventional formative and summative assessment and the LOA-integrated flipped classroom approach yielded a positive effect on TEVT students' language proficiency.

Table 5. Paired samples test

		Paired differences					t	df	Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
Control group	Posttest-pretest	19.379	15.657	2.056	23.496	15.262	9.426	57	.000
Experimental group	Posttest-pretest	21.115	13.901	1.780	24.675	17.554	11.863	60	.000

When differences are found between two groups or two sets of data from a single independent sample, it is crucial to calculate the effect size to assess its practical significance and statistical power (Ambarcıoğlu, 2021; Bickman & Rog, 2009). Among the measures for calculating effect size, Cohen's dz value is particularly well suited to paired or within-subjects designs, as it accounts for within-participant variability. (Bartlett, 2019). Cohen's dz value is calculated with the formula

$$d = \frac{M_{diff}}{SD_{diff}}$$

where the mean values of the difference scores are divided by the standard deviation (SD) of the difference (Dankel & Loenneke, 2021); when the mean difference was divided by SD, for the experimental group, Cohen's dz was 1.52 (M_{diff}=21.115, SD_{diff}=13.901). In contrast, for the control group, Cohen's dz was 1.24 (M_{diff}=19.379, SD_{diff}=15.6573). Both demonstrated a large effect size, suggesting substantial pretest-posttest improvements. It could be concluded that both control groups and experimental groups demonstrated significant improvements in language proficiency over the 12 weeks of language learning. These findings suggested that both LOA-embedded and formative -summative assessment-embedded flipped classroom approaches positively influenced students' language proficiency. However, it remained uncertain what might be held accountable for the improvement.

Independent sample t-test

An independent-samples t-test was then conducted to determine whether the intervention was responsible for the improvement. To this end, a comparison of pretest mean scores between the control and experimental groups was conducted to determine whether any significant differences existed before the intervention. Given the absence of a significant pretest difference between the experimental and control groups, any significant posttest difference can reasonably be interpreted as evidence of an intervention effect. If significance was found in pretest scores and baseline equivalence could not be assumed, analysis of covariance (ANCOVA) should be used to examine posttest outcomes, with pretest scores as a covariate to control for baseline differences.

As shown in Table 6, Levene's test for equality of variances for pretests (Sig. = 0.326) and posttests (Sig. = 0.271) yielded p-values greater than 0.05, indicating that equal variances can be assumed. Thus, the assumption of homogeneity of variances was met, indicating that the two groups exhibited comparable variability in their test scores. Levene's test result, assuming equal variance, revealed that the significance values from the two-tailed tests for the pretest (p = 0.427) indicated no significant difference between the groups. Accordingly, posttest scores were compared using an independent-samples t-test. The results, however, showed that the difference between the experimental and control groups was not statistically significant (p = 0.930). This indicated that when the experimental and control groups began with comparable levels of language proficiency, no statistically significant difference (p = 0.930) was observed between their test scores after 12 weeks, even though the experimental group showed a slight improvement in mean posttest scores compared to the control group.

Table 6. Independent samples test

		Levene's test for equality of variances		t-test for equality of means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
									Lower	Upper
Pretest	Equal variances assumed	0.971	0.326	0.796	117	0.427	1.4836	1.863	-2.206	5.174
Posttest	Equal variances assumed	1.225	.271	-0.088	117	0.930	-0.2518	2.853	-5.901	5.398

Also, the effect size was calculated to demonstrate the magnitude of the difference between groups. Eta squared (η^2) was used to measure the variance of the dependent variable, language proficiency, which was explained by an independent variable or factor, LOA intervention (Yagin et al., 2024). The formula used is as follows: The magnitude of the differences in the mean for pretest scores [mean difference = 1.483, 95% confidence interval (5.1737, -2.2064)] was very small (eta square = 0.005), while the magnitude of the difference in the mean for posttest scores [mean difference = 0.251, confidential interval (5.3979, -5.9016)] was even smaller (eta square = 0.0001). This negligible difference between groups further supports the conclusion that the intervention did not have a meaningful effect. However, considering the experimental group's lower pre-intervention mean compared to the control group, followed by its higher post-intervention mean, the experimental group demonstrated some progress and showed potential to significantly outperform the control group with a longer intervention.

Effect of LOA Intervention on Sub-Skills

A closer examination of student performance in the PRETCO- A test before and after the intervention revealed that, for the experimental group, the posttest scores exceeded the pretest scores in all subsections – listening comprehension, mastery of vocabulary and grammar, reading, and writing – with the sole exception of translation skills. The results of the pretest–posttest paired-samples tests for the five skills in both the experimental and control groups are presented in Tables 7 and 8.

As shown in Tables 7 and 8, the significance value of the pretest-posttest difference for experimental groups in each section is $0.000<0.05$, indicating a significant difference before and after intervention, while in translation skills, the value is 0.085, above 0.05. In the same vein, the significance value for the control group in translation skills is even larger, at 0.289, far above 0.05, indicating a minor enhancement in translation skills.

Table 7. Pretest-posttest paired samples test for sub-skills measured (experimental group)

		Paired differences					t	df	Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
Pair 1	Listening1-Listening2	-4.39	4.92	0.63	-5.65	-3.13	-6.97	60	0.000
Pair 2	Vocabulary1-Vocabulary2	-7.74	7.98	1.02	-9.78	-5.69	-7.57	60	0.000
Pair 3	Reading1-Reading2	-4.20	6.50	0.83	-5.86	-2.53	-5.04	60	0.000
Pair 4	Translation1-Translation2	-0.690	3.07	0.39	-1.47	0.10	-1.75	60	0.085
Pair 5	Writing1-Writing2	-4.098	3.341	0.428	-4.954	-3.243	-9.58	60	0.000

Table 8. Pretest-posttest paired samples test for sub-skills measured (control group)

		Paired differences					t	df	Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
Pair 1	Listening1 - Listening2	-4.172	3.807	0.4999	-5.173	-3.171	-8.346	57	0.000
Pair 2	Vocabulary1 - Vocabulary2	-7.379	8.243	1.082	-9.547	-5.212	-6.817	57	0.000
Pair 3	Reading1 - Reading2	-4.500	7.712	1.013	-6.528	-2.472	-4.444	57	0.000
Pair 4	Translation1 - Translation2	-0.466	3.315	0.435	-1.337	0.406	-1.069	57	0.289
Pair 5	Writing1-Writing2	-2.86	3.05	0.401	-3.664	-2.060	-7.142	57	0.000

To examine the scores of each language skill measured closely between the experimental group and the control group, statistical analysis revealed that the scores of the experimental and control groups did not differ significantly on any of the five skills in the pretests. The two-tailed tests, assuming equal variances, yielded non-significant p-values for listening (P = 0.612), vocabulary and grammar (P = 0.093), reading (P = 0.888), translation (P = 0.509), and writing (P = 0.913), indicated in Table 9.

Similarly, an independent sample test was conducted on posttest scores. The results also indicated that no significant differences were found between the experimental and control groups in listening ($p = 0.911$), vocabulary and grammar ($p = 0.466$), reading ($p = 0.715$), and translation ($p = 0.226$), except for writing ($p = 0.039 < 0.05$). Given that there was no significant difference between the experimental and control groups in pretest writing scores ($p = 0.913$), the significant difference in writing outcomes can be attributed to the intervention, suggesting that integrating LOA within a flipped classroom may improve students' writing skills. The eta squared (η^2) was calculated as approximately 0.0363, demonstrating that 3.6% of the variance in the writing skills can be explained by the LOA integrated flipped classroom framework. It is a small effect size, given that $\eta^2 = 0.01$ indicates a small effect, $\eta^2 = 0.06$ indicates a medium effect, and $\eta^2 = 0.14$ indicates a large effect (Yagin et al., 2024).

Table 9. Independent samples test for sub-skills measured (pretest)

		Levene's test for equality of variances		T-test for equality of means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
									Lower	Upper
Listening (Equal variances assumed)	pretest	0.127	0.722	0.509	117	0.612	0.312	0.613	-0.903	1.527
	posttest	0.072	0.788	0.112	117	0.911	0.091	0.809	-1.511	1.693
Vocabulary (Equal variances assumed)	pretest	0.169	0.682	1.696	117	0.093	1.360	0.802	-0.228	2.947
	posttest	2.27	0.135	0.732	117	0.466	1.001	1.368	-1.708	3.710
Reading (Equal variances assumed)	pretest	0.053	0.818	0.141	117	0.888	0.127	0.901	-1.657	1.911
	posttest	2.60	0.110	0.366	117	0.715	0.430	1.175	-1.898	2.757
Translation (Equal variances assumed)	pretest	2.55	0.113	-0.662	117	0.509	-0.280	0.423	-1.117	0.557
	posttest	0.141	0.708	-1.22	117	0.226	-0.503	0.413	-1.321	0.316
Writing (Equal variances assumed)	pretest	0.328	0.568	-0.110	117	0.913	-0.0348	0.3158	-0.6602	0.5907
	posttest	1.06	0.306	-2.09	117	0.039	-1.271	0.609	-2.477	-0.065

Effective LOA Strategies

A focus group was conducted after the quasi-experiment, with both teacher participants and expert observers participating. The discussion centered on the following three questions: (1) Which technology-assisted LOA strategies do students find most beneficial for improving their learning outcomes?

(2) What challenges do students and teachers encounter when implementing LOA practices in technology-enhanced environments? (3) How are technologies incorporated to promote the LOA process?

Findings from the focus group discussion validated the operationalization of three core LOA principles: assessment tasks for learning purposes, active student participation in assessment, and the strategic application of feedback as feedforward oriented towards future learning. Among the recommended strategies, educators recognized rubric-driven peer assessment as a key component of a successful learning approach, noting that structured peer assessment activities can positively promote deep learning. As another teacher noted, "peer assessment, when structured effectively with tools like rubrics, had the potential to transform the conventional assessment process into a powerful learning experience." During peer assessment, students perceived a critical need to engage deeply with the learning content and meticulously review the assessment rubrics before they could generate high-quality feedback. One teacher participant noted that "after peer assessment, students gained a better understanding of the material." To do this, one teacher suggested that "before doing self-assessments or peer assessments, teachers should clearly explain the rubrics". This preparatory phase served to intensify their interaction with the course material and to foster their evaluative expertise.

Another workable strategy was to offer students a chance to give feedback to their peers after peer review and assessment. For example, in an English text-based mind map assessment activity, it was observed that students who participated in peer review and provided constructive feedback simultaneously achieved two distinct advantages. First, peer feedback activities helped sharpen students' analytical skills by requiring them to examine others' work critically. This critical examination typically involved pinpointing both the shortcomings and merits of a submission, specifically concerning rubrics such as the accuracy of keyword choices, the explicitness of relational connections, and the completeness of the presented content. Second, it promoted engagement with learning content and peer learning. Concurrently, the teachers noted that by emulating established patterns of English expression observed in proficient English speakers, students can enhance their own written work and broaden their understanding. These enabled students not only to evaluate their peers' work fairly but also to engage in a deeper, more collaborative form of learning. To incentivize students' involvement and engagement in peer assessment and feedback, teachers factor students' performance in peer assessment and the quality of peer feedback into the final course grade.

Moreover, teacher participants also highlighted action-oriented feedback loop integrating individualized e-feedback and teacher guidance as a key LOA strategy, emphasizing its role in supporting learners' reflection and ongoing improvement. The teacher participants suggested that feedback should be given in detail and in written form, including the learning objectives or content, diagnosis of the learning gap, and required follow-up actions. Table 10 presents the feedback from teachers following the unit tests. The feedback enables students to identify areas for improvement by comparing their individual results with the class means.

Table 10. Template of learning feedback delivered via u-learning platform

Learning Feedback for Weeks_____

Posted by: _____ Date: _____ Class: _____ Read by:_____

Learning Content:

Unit __	Vocabulary	
Unit __	Text reading	
Unit __	Translation	
Unit __	Listening &speaking	
Unit __	Grammar	
Unit __	Writing	

Learning Outcomes:

1. **Unit test results**:

Average score: _____ Highest score: _____ Lowest score: _____Students at risk: _____

Scores between 60–70: _____ Scores between 70–80: ____ Scores between 80–90:_____

Breakdown of performance by question type:

Section	**Average**	**Highest**	**Lowest**	**Std. Dev.**
Vocabulary MCQs				
Grammar MCQs				
Listening comprehension				
Word formation cloze				
Reading comprehension				
Translation (MCQs &cloze)				
Writing task				
Oral test				

Areas to Improve:

From the table above, scores for _____were relatively low. This indicates that some students still need to improve their _____ and_____. Vocabulary/ grammar such as _____ as well as _____(sentence structure) remain areas to work on.

2. **Writing Feedback**:

Content & task fulfillment:	_____
Language accuracy:	_____
Tone& Style:	_____
Structure & coherence:	_____

Required Follow-Up Actions:

☐ Watch the video tutorial on how to make corrections.

☐ Revise and resubmit writing task in accordance with the feedback.

☐ Complete the self-reflection sheet for the unit.

☐ Create a self-reflection report to examine your learning progress and learning gap

☐ Complete targeted online practice

Encouragement:

I hope everyone keeps up their enthusiasm, continues to work hard, and achieves even better results!

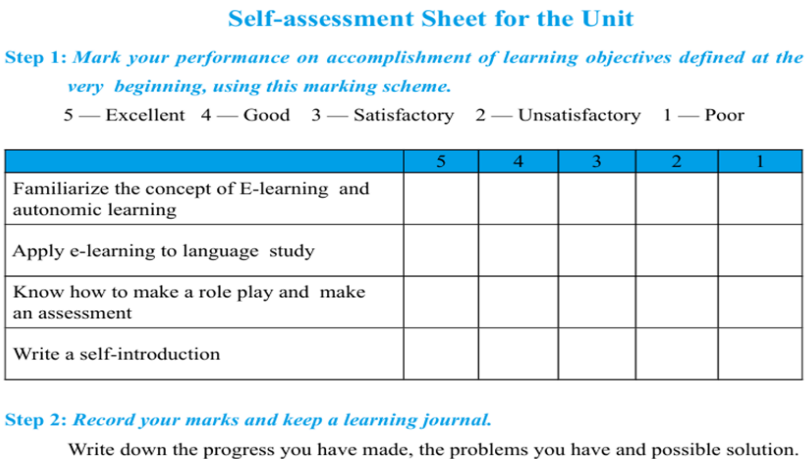
In addition to the feedback from teachers to the whole class, the automatic scoring system embedded in the LMS also provided instant feedback on students' writing, suggestions for improvement, as well as detailed profiles on individual students' learning outcomes, which help both teachers and students make informed decisions on instructional and learning strategies. However, while e-assessment and

feedback offer the advantage of timeliness, they may lack informativeness and accuracy, particularly with respect to suggestions for improving written work. The teacher participants reported that they were more likely to use these tools in complementary roles alongside their own feedback and calibration, rather than as a standalone method. Additionally, they highlighted the importance of tracking students' responses to feedback, specifically whether they had reviewed, reflected on, and revised their written work in accordance with it. To ensure this, teachers required students to resubmit their revised work, and scores for the final versions were factored into the course grade. During the resubmission and revision process, feedback serves as instruction and scaffolding, continually helping students improve their learning. It also provides them with an opportunity to demonstrate their comprehension of the feedback and their overall language acquisition. As another teacher concluded, "when students receive immediate feedback on their pre-class work, they feel their pre-work is valued and develop a sense of recognition". This positive reinforcement, whether through praise or constructive guidance, is key to increasing their engagement and preventing disengagement. This establishes a self-sustaining learning loop: preparation leads to assessment, assessment to feedback, and feedback to improvement.

Teacher participants identified self-assessment and self-reflection informed by feedback and aligned with the desired learning outcome as the third effective strategy. As one teacher mentioned, this approach empowered students to actively participate in the evaluation process through well-structured self-assessment rubrics (as shown in Figure 5) and reflective journals. For instance, at the conclusion of each unit, students were required to assess their performance in key areas, such as thematic comprehension, vocabulary, and communication skills, against the desired learning outcome, while simultaneously identifying topics for remediation. It was observed, by reviewing their journal entries, students then independently synthesized their own effective and less effective learning strategies. The teacher believed that "ultimately, this process positions students as autonomous managers of their learning, shifting them from passive recipients to active participants." This aligns with the educational objectives of fostering students' self-directed learning skills.

Strategic Incorporation of Technology Into LOA

In this research, the U-learning platform (Ubiquitous Learning Platform), which integrated learning resources delivery, learning management, and assessment, was used to enhance LOA practice in the flipped classroom context. The smart education platform, developed by Beijing Wenhua Online Education Technology Co Ltd, especially for English language teaching, learning, and assessment, incorporates advanced educational technology and enables assessment of learning, assessment for learning, assessment as learning, and the seamless integration of LOA into flipped classroom instruction.



Self-assessment Sheet for the Unit

Step 1: ***Mark your performance on accomplishment of learning objectives defined at the very beginning, using this marking scheme.***

5 — Excellent 4 — Good 3 — Satisfactory 2 — Unsatisfactory 1 — Poor

	5	4	3	2	1
Familiarize the concept of E-learning and autonomic learning					
Apply e-learning to language study					
Know how to make a role play and make an assessment					
Write a self-introduction					

Step 2: ***Record your marks and keep a learning journal.***

Write down the progress you have made, the problems you have and possible solution.

Figure 5. Screenshot of the self-assessment sheet for the unit

Technology-supported LOA practice

According to the focus group discussion, technology played an irreplaceable role in promoting LOA practice in this research. It enabled instructors to collect impromptu learning evidence for assessment for learning, increased the authenticity and efficiency of assessment for learning and feedback, and provided opportunities for students to assess their peers' work using embedded rubrics. As shown in Table 11, the focus group discussion highlighted the enabling role of the U-learning platform in supporting the basic elements of LOA, such as assessment of learning, assessment for learning, and assessment as learning.

Table 11. Technology support of the U-learning platform for LOA practice

LOA elements	Module	Major features	Technical support
Assessment for learning	Information delivery	Class-specific delivery of online courseware resources, personalized learning pathway push, specification and release of learning tasks, and feedback provision to targeted classes.	Content management, adaptive learning algorithms, push services, workflow & task management, database management.
	Learning process monitoring	Tracking learning progress, fraud detection, grade calculation, and presentation.	Learning analytics; AI algorithms; event logging & activity tracking, fraud detection algorithms, weighted scoring, data analysis, and visualization.
	Assignment management	Assignment submission via images, audio, and text input, automatic grading for writing and speech, e-feedback, batch manual grading, and attempt management.	Multi-format upload support, cloud storage, natural language processing, speech recognition/ automatic speech scoring, automated commenting & annotation, grading management, version control, workflow status, and inline annotation.
	Interactive tools and synchronous interactive assessment	Polls, quizzes, online discussions, screen casting, real-time comments, random selection, quiz bowl, AI-generated questions, photographs, and an uploading and scoring system.	Real-time communication protocols, gamification modules, NLP algorithms, randomization algorithms, question banks, camera integration, and real-time event synchronization.
	Data analytics and presentation	Automated learning analytics, learning dashboard, screen cast, and instant performance reporting.	Multimodal data integration, multimedia display integration, intelligent diagnostic analysis, real-time data sync, automated reporting, data analytics & visualization.

LOA elements	Module	Major features	Technical support
Assessment of learning	Test bank and paper generation	Collaborative item bank construction, automated exam paper creation, randomized question selection, timely paper generation, and difficulty-level classification of test items.	Content management system, collaboration tools, algorithmic test assembly, assessment platform integration, template & formatting tools, statistical analysis & item response theory, and metadata tagging.
	Mobile examination and e-proctoring	Item shuffling, randomized exam setting, online exam proctoring, abnormal exam activity alerts, automatic grade calculation, results generation, and reporting.	Exam management, UI/UX design, AI-based proctoring tools, screen-switch monitoring, IP & device monitoring, AI anomaly detection, weighted scoring systems, data analytics, and report automation.
	Exam scoring and feedback	Automatic writing evaluation (AWE) and e-feedback, auto-scoring tools, AI-generated feedback, manual grading, and teacher feedback.	NLP, speech recognition & scoring engines, rule-based algorithms, machine learning models, grading management, rubric-based digital tools, annotation & markup tools, language generation models, and feedback modules.
Assessment as learning	Peer assessment	Rubric-guided assessment, random allocation of assessors, peer feedback, and appeal.	Digital rubric management, randomization algorithms, workflow automation engines, peer review management, anonymization tools, collaboration tools, appeal management, and version control.
	Group activities	Automatic grouping and manual grouping, individual or collective grading, and online group discussion	Clustering algorithms, instructor control interfaces, grading engines with flexible logic, audit trails, discussion forums, and content sharing tools.

First, the U-learning platform was architected to facilitate LOA by embedding assessment directly into the continuous learning cycle and interactive instruction. Beyond content delivery, it supported the implementation of continuous, low-stakes assessment; the delivery of instantaneous feedback (feed-forward); the capture of spontaneous or impromptu classroom-based assessment data, and the recording of follow-up actions, thereby making for a performance-interpretation-feedback-instruction cycle. Moreover, the platform capitalized on interactive technologies, such as real-time communication protocols, a gamification module, and randomization and Natural Language Processing (NLP) algorithms, to enable broader student engagement in the classroom interaction and track how

their interlanguage evolved through interactions with peers, digital learning content, embedded rubrics, and direct teacher feedback.

In addition, by recording students' interactions with online resources, the platform generated granular learning trajectory data, which informed teachers' pedagogical adjustments and triggered automated alerts for anomalous performance patterns (e.g., a sustained decline in pronunciation scores). Moreover, the learning analytics technology was leveraged to analyze engagement and performance data, enabling it to proactively identify students exhibiting a "high engagement–low mastery" profile. This diagnostic capability allowed for immediate and targeted instructional intervention, facilitating the instructor-guided learning. The dynamic capability supported the design of differentiated assignments and adaptive learning pathways, fully aligning the platform's functionality with the core LOA principle of using evidence to advance student learning. Doing these helps improve the efficiency of evidence collection for learning and the validity of assessment for learning.

Second, the platform also enhanced learning assessment by employing a dual system of e-proctoring and data analysis. To ensure the authenticity and reliability of assessment data (Assessment of Learning), the platform implemented robust anti-fraud measures, including randomized test generation and a comprehensive online proctoring system. This proctoring framework utilized multi-layered monitoring, including device and IP surveillance, screen-switching detection, facial recognition, and AI-based posture analysis, to identify and document irregularities. This strict control ensured that the resulting performance metrics provided reliable evidence of student online learning outcomes. Meanwhile, the test system allowed teachers to randomize questions from the item banks to generate individualized exams for each student, thereby minimizing the risk of peer-to-peer cheating. By yielding highly reliable data, the system allowed instructors to leverage AI technology for pedagogical improvement, specifically to diversify and refine assessment task design and generate high-quality, actionable feedback, effectively supporting the crucial "feed-forward" principle of continuous improvement.

Third, the platform significantly advanced LOA by implementing a three-dimensional evaluation matrix that fostered self-regulation and evaluative expertise. This matrix systematically integrates automated scoring, peer review, and instructor evaluation to maximize both the reliability of performance data and the transparency of assessment criteria. This integration immediately diversified assessment modalities and improved efficiency. The platform's automated capabilities generated diagnostic indicators (e.g., accuracy rates and oral fluency) instantly, significantly reducing teacher workload. The rubric-driven peer assessment and feedback mechanisms fostered collaborative learning and peer learning among students.

Furthermore, the visual representation of answer distributions, class ranking, and points awarded facilitated students' interpretation of performance data and reflection on their learning gaps. By actively engaging students in the critical evaluation of their peers' and their own work, the platform directly supported the development of students' evaluative expertise, the necessary skills for self-regulation. It informed judgment of quality, thereby enriching the overall learning environment. Overall, the use of technology in LOA plays a positive role in enhancing efficiency, authenticity, and reliability of assessment of learning, sustaining continuous feedback cycles, and supporting data-informed instructional decision-making.

Implementation challenges

Despite those effective strategies, teacher participants also noted some implementation challenges that might limit the impact of LOA. The primary challenge lay in designing assessment tasks. Teachers found it challenging to align assessment and learning tasks with the prescribed curriculum objectives while simultaneously addressing a fundamental issue: students' poor foundational language skills. In China, the rigid, prescribed nature of the curriculum objectives and learning content presented a significant challenge for English language teachers. As one teacher participant mentioned,

"teachers must design assessment tasks that are appropriately calibrated for difficulty, taking into account students' prior knowledge, while simultaneously ensuring these tasks accurately reflect the learning objectives mandated by the curriculum standards." Moreover, ensuring consistency in assessment difficulty and alignment between assessment tasks and instructional focus posed a significant, multifaceted challenge for educators. The teacher participants noticed that despite the high average unit test scores, there had been no significant improvement in student performance from one unit to the next. They reckoned that this lack of progress might be due to two primary issues: inconsistent difficulty in assessment tasks and poor alignment between assessment tasks and instructional focus. As one teacher said, "the unit tests mainly focus on difficult concepts, but they might not cover everything students need to know." The gap between what students have learnt and what they are tested on, which, according to the teacher participants, might lead to students' disengagement and frustration. For instance, they noticed that "pre-class assessment tasks lacked relevance and an appropriate level of challenge, which diminished student engagement and discouraged them from completing the assignments." Consequently, these assessment tasks failed to promote skill development effectively. Finally, the teachers also emphasized the importance of designing learning tasks with a real-world context, allowing students to apply their knowledge in authentic situations. However, this places high demands on a teacher's English language instructional skills and LOA literacy, requiring specialized training in crafting learning-oriented assessment tasks.

Another significant challenge in LOA implementation is academic dishonesty resulting from students' misuse of technology. The discussion indicated that although technology-enhanced classroom interaction enabled alternative forms of assessment and provided instant feedback, its misuse in LOA also posed risks to the authenticity of assessment outcomes, particularly in activities conducted beyond the classroom without direct teacher supervision and proctoring. This could yield distorted information about students' learning gaps and foster a misleading impression of their language proficiency and learning progress, ultimately leading to an inappropriate pedagogical response. Teacher participants believed this behavior might stem from a combination of factors, including a lack of student motivation, tight deadlines, and tasks perceived as being overly difficult. Nevertheless, the key contributing factor was the quantitative focus of the assessment results. When LOA assessment was conducted exclusively through an LMS that recorded and quantified activities, this approach, unlike those prioritizing qualitative or descriptive outcomes, could inadvertently encourage academic dishonesty.

Some students were found to abuse technology to quickly meet the numerical requirements, thereby avoiding the cognitive effort required for deeper learning. The teacher cited an example in which TVET students were informed that formative assessments, quantified by the LMS for both in- and out-of-class performance, accounted for 60% of their overall grade, and that they were more actively engaged in classroom interactions. However, this high weighting also incentivized them to perform only the bare minimum required to meet the learning criteria, rather than to engage with the material for deeper comprehension. As one teacher noted, "some students, when completing preview learning tasks, focus on meeting the required quantity of work but pay little attention to its quality or depth." Another teacher posited that "a 60% weighting might not be reasonable, since it could actually encourage students to cheat. We could consider alternative forms of assessment, including traditional paper-and-pencil exams and performance-based tasks such as presentations or role-plays." However, they were also aware that "for large classes, it is more practical to manage formative assessments using technology and quantitative measures. The balance between formative assessment and the final exam, as well as the way results are reported quantitatively, has been established by the academic authority. Making changes to this would mean rethinking the whole assessment approach." So, despite the benefits of LOA to TEVT students, as one teacher summarized, it turned out that "the assessment activities were found to be most beneficial for students who exhibited strong intrinsic motivation, a passion for learning English, and a high level of concentration."

RESPONSES TO THE RESEARCH QUESTIONS

Based on the results of the study, the following sections address the research questions:

RQ1: Significant improvement was observed in experimental groups, indicating that the LOA-integrated flipped classroom model had a positive effect on vocational college students' language learning outcomes with a large effect size ($P = 0.000$, $t = 11.863$, Cohen's $dz \approx 1.52$, $N = 61$). The intervention was incorporated into a flipped classroom instructional model and empowered by educational technology, including automatic scoring of assessment of learning, LMS-enhanced assessment for learning, and assessment as learning. LOA strategies implemented also included learning tasks as assessment tasks, student engagement in assessment, and feedback as feedforward. However, significant improvement was also observed in the control group between pretest and posttest ($P = 0.000$, $t = 9.426$, $N = 58$, Cohen's $dz \approx 1.24$). This finding suggested that incorporating both LOA and conventional summative-formative assessment into flipped classroom instruction improved TEVT students' learning outcomes. However, it cannot be inferred that the observed significant improvement in scores between the pretest and posttest was a direct result of the technology-enhanced LOA intervention.

RQ2: Between-group scores indicated that the intervention demonstrated a positive effect on overall language proficiency, although the difference was not statistically significant. The independent-samples t-tests indicated that there was no statistically significant difference in language proficiency scores between the experimental groups and the control group at the pretest [$t = 0.796$, $P = 0.427$, $df = 117$, 95% confidence interval (-2.206, 5.174), $\eta^2 = 0.005$]. In the posttest, the difference remained insignificant [$t = -0.088$, $P = 0.930$, $df = 117$, 95% confidence interval (-5.902, 5.398), $\eta^2 \approx 0.0001$], only that considering the experimental group's lower pre-intervention mean compared to the control group, followed by its higher post-intervention mean, the experimental group demonstrated notable progress, suggesting a potential for significantly outperforming the control group with a longer intervention. This suggested that the technology-enhanced LOA intervention failed to yield significantly greater effects on TEVT students' language proficiency than conventional formative plus summative assessment practices.

RQ3: Despite the significant differences before and after intervention for experimental groups across the language skills ($p = 0.000 < 0.05$) except translation skills ($p = 0.085 > 0.05$), no significant difference was found between experimental groups and control groups across the sub-skills measured (listening $p = 0.911$, vocabulary $p = 0.466$, reading $p = 0.715$, translation $p = 0.226$) except writing skills ($p = 0.039$) with a small effect size ($\eta^2 \approx 0.0363$). Thus, it can be concluded that the LOA intervention had a limited effect on TVET students' listening, vocabulary and grammar, reading, and translation skills, but demonstrated a significantly positive effect on writing skills.

RQ4: Based on the findings from focus group discussions, effective LOA strategies included rubric-driven peer assessment activities bolstered by preparatory rubric-based training, and an action-oriented feedback loop integrating individualized e-feedback and teacher guidance, protocol-based self-assessment, and self-reflection informed by feedback and aligned with desired learning outcomes. These strategies facilitated deep engagement with learning content and peer learning, sharpened students' critical and analytical skills, and positively affected students' language learning.

RQ5: The findings from the focus discussion indicated that technology was capitalized to support assessment for learning, assessment of learning, and assessment as learning. It enhanced multimodal interactions, improved the efficiency of continuous, low-stakes assessments, delivered instantaneous feedback, and captured spontaneous assessment data, thereby greatly improving the time efficiency and validity of assessment-driven learning.

The dual system of e-proctor and performance data analysis ensured the authenticity and reliability of the assessment of learning. The rubric-driven peer assessment and feedback mechanisms fostered peer learning and self-reflection. However, it also raised concerns about academic dishonesty and abuse of technology.

DISCUSSION

This research investigated the effect of LOA on TEVT students' English language proficiency, synthesized the LOA strategies that were yielding positive outcomes, and analyzed the role of technology in operationalizing LOA. Regarding the first and second research questions, which examined the effect of LOA on TEVT students' language proficiency, the findings indicate that students exposed to LOA intervention demonstrated no significant improvement in comparison to those in a conventional assessment context, despite significant gains before and after the intervention. Regarding the third research question as to what specific language skills were affected positively by technology-enhanced LOA intervention, the data analysis revealed that among all the language skills and knowledge being measured, only writing skills were found to be significantly improved. The examination of effective LOA strategies suggested that rubric-driven peer assessment, action-oriented feedback loops, and protocol-based self-assessment and reflection might contribute to the improvement of writing skills and self-assessment. Finally, the investigation into the role of technology revealed that while technologies enhanced LOA and promoted the integration of LOA into flipped instruction, they also introduced concerns regarding academic integrity and potential erosion of deep cognitive engagement.

EFFECTS OF LOA ON THE ENGLISH LANGUAGE PROFICIENCY

RQ1. Is there a notable difference in the English language proficiency levels of vocational college students before and after technology-enhanced LOA treatment?

The within-group comparison indicated that the technology-enhanced LOA in a flipped classroom setting demonstrated a positive effect ($P < 0.001$, $t = 11.863$, $dz = 1.52$, $N = 61$) on TEVT students' language proficiency, with a large effect size (Cohen's $dz \approx 1.52$), suggesting substantial efficacy. To interpret the findings, a comparison of the technological integration and assessment protocols implemented in the control and experimental groups (Table 1) reveals that the two groups share an interactive learning environment supported by interactive tools, with a blend of product- and process-oriented assessments. These interactive tools and assessment mechanisms with a dual focus on learning outcomes and learning processes fostered students' active classroom participation and learning motivation, which might account for the observed gains in vocational college students' language learning (Waskito et al., 2021). This aligns with Active Learning Theory, which posits that effective learning arises from active engagement in the learning process rather than passive information reception (Junejo et al., 2022).

For flipped classroom practitioners and educators, this suggests that the incorporation of interactive tools into process-focused formative assessment could produce an interactive learning experience, promoting the transformation of TEVT students in China from passive learners to active learners. More importantly, the integration of the two drives the pedagogical process to ensure students stay cognitively engaged. The focus of teacher training should be on maximizing the potential of interactive tools to address the challenges in learning and assessment in large classes and on how to refine the design of assessment tasks with the assistance of technology to create evidence-based learning experiences to foster maximal engagement.

RQ2. Is there a notable difference in the English language proficiency levels between students in the intervention and control classes before and after the intervention?

As mentioned, the independent sample T test results indicated that intervention groups did not significantly outperform control groups. However, the absence of observable behavioral change does not in itself invalidate LOA, since, as Gebril (2021) argues, assessment tasks designed for learning-oriented purposes may at times fail to yield immediate or measurable learner improvement. It is confirmed by Viengsang and Wasanasomsithi (2022), whose research found that the LOA model had no statistically significant effect on students' reading skills.

To justify the insignificant effects, as mentioned, Estaji and Safari (2023) found significant differences between reflective learners and impulsive learners, suggesting varying benefits of the LOA intervention across different learners. Wang (2022) echoed that LOA might have a diversified outcome among participants. In this research, the teacher participants also implied that assessment activities proved most effective for students who already possessed specific learning attributes. These students were intrinsically motivated, meaning their drive came from a genuine interest in English rather than from external rewards or pressure. This internal passion for the subject, coupled with a high level of concentration, allowed them to engage more deeply in the assessment tasks. Thus, it can be inferred that the possible reason for the LOA intervention yielding an insignificant effect might be that some students were not intrinsically motivated despite technology-enhanced LOA integration. According to Self-Determination Theory (SDT), the satisfaction of students' three basic psychological needs, namely competence, autonomy, and relatedness, could yield intrinsic motivation (Ryan & Deci, 2000). The LOA strategies, such as timely feedback, feedback loop, and peer assessment, supported these three needs and were expected to arouse students' intrinsic motivation, thereby improving learning outcomes (McMillan & Hearn, 2008). However, as shown in Table 10, the formative feedback was mostly quantitative and grade-based, and the students received mostly extrinsic rewards for learning efforts rather than constructive feedback, which might lead to superficial engagement in learning.

Thus, for teachers, it suggests that hybrid feedback, including both instant formative feedback provided by LMS and delayed qualitative feedback, could be more effective than formative feedback alone (Aliakbari et al., 2025). It also serves as a pivotal reminder for teachers that a narrow focus on grade-based assessment cannot warrant sustained motivation and engagement. Thus, institutions should shift from quantitative grade reporting to detailed descriptive reports on the profile of students' interlanguage development.

RQ3. What's the effectiveness of technology-enhanced LOA on specific language skills?

Although the overall pretest–posttest differences in total scores were not statistically significant, an analysis of specific language skills and linguistic knowledge revealed a significant improvement in writing scores after the intervention ($p = 0.039$, $t = -2.09$, $df = 117$). This suggests that the integration of LOA exerts a positive influence on students' writing abilities. More than one study has verified the effectiveness of LOA in improving students' writing skills (e.g., Imsa-ard, 2024; X. Yang, 2020). These findings suggest that LOA has a positive effect on the writing skills of foreign language learners, regardless of their initial proficiency level; however, the extent and nature of the benefit may vary among different learner types. The limited improvement in other sub-skills suggests that the LOA intervention has differential effects on various aspects of language proficiency. This conclusion is consistent with studies by Y. Zhao et al. (2024), who found that the intervention had a limited impact on students' overall English proficiency comprehensively, but a significant effect on one specific skill. These divergent outcomes suggest that different facets of language proficiency may respond differently to the pedagogical interventions.

Several factors may account for the statistically insignificant difference observed in language skills other than writing. From the lens of zones of proximal development (ZPD), qualitative exploration

on the challenges of LOA practice revealed that the misalignment between mandatory learning content and standardized unit tests, together with the gap between students' differentiated academic backgrounds and uniform curriculum objectives, compromised student engagement by placing learning demands outside many learners' ZPD (Vygotsky, 1978).

From the LOA framework, assessment tasks that lack relevance to real-world contexts further diminish students' interest in learning. In this research, to incentivize students' engagement with feedback, grades on writing works were assigned not to the initial drafts but to the final submissions following multiple rounds of revision. By incentivizing students' investments in review and revision, this approach ensures that students' efforts in revising their work and their understanding of the feedback are considered, thereby encouraging critical reflection on their learning gaps and the implementation of corrective actions (McBeth, 2015). Also, as mentioned above, although the incentive measures, such as extrinsic rewards, may foster a sense of progress, mastery, and competence in line with Self-Determination Theory (Ryan & Vansteenkiste, 2023), they may also shift students' focus on grades rather than learning processes, thereby diverging from the core principles of LOA.

Thus, it remains a compelling challenge for teachers to determine how to motivate students to engage with feedback and assessment tasks without perpetuating a grade-centric mindset. Concurrently, educational institutions and administrative bodies should consider granting teachers greater pedagogical autonomy, allowing them to recalibrate curricular content and objectives to reflect the TEVT students' English interlanguage development and the needs of vocational English learning in a globalization context.

LOA Strategies in a Flipped English Classroom Setting

RQ4. Which specific LOA strategies are effective in improving students' language proficiency?

Qualitative exploration found that rubric-driven peer assessment, action-oriented feedback loops, and protocol-based self-assessment and self-reflection might all contribute to positive learning outcomes. Among all these strategies, feedback was recognized as a key element of effective classroom assessment practices (Zúñiga, 2024). In research by Estaji and Safari (2023), the participants in both groups also recognized the positive role of teachers' feedback and peer assessment in enhancing their writing skills. Furthermore, Carless (2015) and Gramcheva (2020) emphasized that it was student engagement with feedback, rather than the mere provision of feedback, that was pivotal for fostering deep learning and enhancing learning outcomes. Table 10 indicates that learning feedback, followed by unit tests, highlights the iterative and reflective process of revision and resubmission for writing tasks. This practice is supported by Estaji and Safari (2023), who claimed that through systematic, repeated cycles of assessment and reassessment, students received feedback that drove them to diagnose their own weaknesses, develop and effectuate corrective measures, and subsequently present a refined demonstration of their knowledge. Kunschak (2020) posited that this closing feedback loop is necessary for avoiding work done perfunctorily and superficially. This feedback loop, characterized by iterative and reflective cycles of revision and resubmission, may help explain the significant gains observed in writing skills. In addition, as aforementioned, hybrid feedback integrating both e-feedback and teachers' qualitative feedback was also recommended (Aliakbari et al., 2025).

Additionally, it has been found that rubric-driven peer assessment can also benefit language learners. Prior to peer assessment, students were trained on how to render valid assessments as well as constructive feedback by studying assessment rubrics embedded in the LMS along with exemplary work. The practices are consistent with the research by Mumpuni et al. (2025), who found that technology-assisted peer assessment, when integrated with elements such as rubric comprehension, quality feedback, and self-reflection, can help improve learning outcomes. However, according to a qualitative study, the practice was constrained by limited class time. Teachers struggle to balance the time allocated in class for peer assessor training versus language learning activities. In response to this

problem, teacher participants tried utilizing the LMS to transition the peer assessor training to short, out-of-class videos.

Finally, the focus group discussion revealed that well-structured rubrics and protocols can effectively guide students through a process of self-assessment and reflection. This is supported by Krebs et al. (2022), who posited that the use of rubrics could not only enhance accuracy but also reduce cognitive load in self-assessment. In this research, a rubric-embedded protocol was developed and implemented at the conclusion of each unit, following the unit tests. It helped students identify their learning gaps and reflect on their performance against predefined learning outcomes. Despite its limited benefit to language skills, protocol-based self-reflection assisted students in comprehending their learning goals, monitoring learning progress, and developing self-directed learning skills (Karaman, 2024).

By capitalizing on technology to scale peer assessment and self-assessment, the study offers a robust roadmap for educators to foster students' evaluative expertise and self-directed learning skills in high-enrollment courses within the LOA framework.

Role of Technologies in LOA

RQ5. What role does technology play in the implementation of learning-oriented assessment (LOA)?

In this research, the U-learning platform, an intelligent LMS, integrates course delivery, learning process monitoring and management, e-testing and e-feedback, and a peer assessment module, demonstrating great potential to support a flipped classroom and enhance learning-oriented assessment. However, it raises concerns regarding student disengagement and academic dishonesty, which might explain the limited impact of LOA on learning outcomes.

It was found that the use of cutting-edge technology in LOA helped to address the challenges found in previous studies. Quyen and Khairani (2017) posited that the difficulty with LOA lay in the implementation, especially in large classes where assessment can be time-consuming and labor-intensive for teachers. This research employed e-testing systems, in which such features as item banking, random construction of tests, data analysis of test results, batch grading, and automatic scoring of writing tests and speeches were embedded, which allowed language teachers to not only administer conventional test types but also performance-based tasks such as situational writing and speaking tests. Moreover, automated scoring systems and AI-generated feedback could serve as valuable tools to enhance the efficiency and responsiveness of assessment practices (Langove & Khan, 2024). The use of integrated platforms promoted formative feedback, adaptive learning, and more authentic assessment practices in language education (S. Li, 2020; Zheng & Yu, 2018), making LOA implementation more flexible and time-efficient (Meccawy et al., 2023).

Another challenge, as Saville (2021) identified, is that the evolution of assessment primarily stemmed from the difficulty of capturing and utilizing sufficient data that definitively evidences the process and outcome of learning. This difficulty is especially pronounced when gathering data from unplanned and impromptu assessment types. Nevertheless, the integration of learning analytics, sophisticated Artificial Intelligence (AI) algorithms, real-time communication protocols, and dynamic item banking served as a foundational infrastructure to mitigate these barriers. Functionalities such as event logging, granular activity tracking, real-time data analysis, and automated reporting afforded substantial epistemic advantages. Collectively, these tools enabled instructors to deploy formative assessment instruments instantaneously during class and acquire timely feedback on student performance. This immediate feedback loop critically enhanced the potential for responsive, evidence-based pedagogical adjustments.

In addition, in this research, technology was capitalized to unlock the potential of LOA for advancing flipped English classroom practice. As shown in Figure 6, before class, the system created a

supportive environment for self-directed learning through a multimodal presentation of linguistic knowledge. Beyond this, it leveraged human-machine interaction to enable students to engage in language practice and obtain instant feedback via speech recognition, NLP, and an automatic scoring system. This instant corrective feedback is especially important for language learning, as it allows students to notice the gap between their current interlanguage and the target language, making it easier to bridge that gap and preventing the fossilization of incorrect forms (Yamamoto, 2003; Yu, 2014). In this sense, the assessment feedback served as scaffolding, facilitating self-reflection and self-directed learning. In the in-class session, technologies such as interactive tools and synchronous assessment modules were integrated into the LMS, enabling teachers to interact with every student simultaneously. This was especially beneficial for students in large classes, as it allowed for individualized attention and fostered meaningful engagement in classroom interaction and assessment activities (Hussain et al., 2019). More importantly, leveraging automated analytics, the LMS synthesized students' learning evidence from classroom interactions into formative feedback to support teachers in interpreting student progress in a time-efficient manner (Martin & Ndoye, 2016; Raghavjee et al., 2021). Coupled with teachers' observation, this data could advance teachers' understanding of students' interlanguage, enhancing their capacity to provide adaptive scaffolding. In this way, the technologies facilitate integrating the LOA cycle into instruction by aligning classroom interaction, timely feedback, and scaffolding, thereby promoting ongoing learning. In the after-class session, the rubric-embedded peer assessment model, AWE system, feedback module, and assignment management and submission system encourage students' engagement in assessment and support the feedback loop.

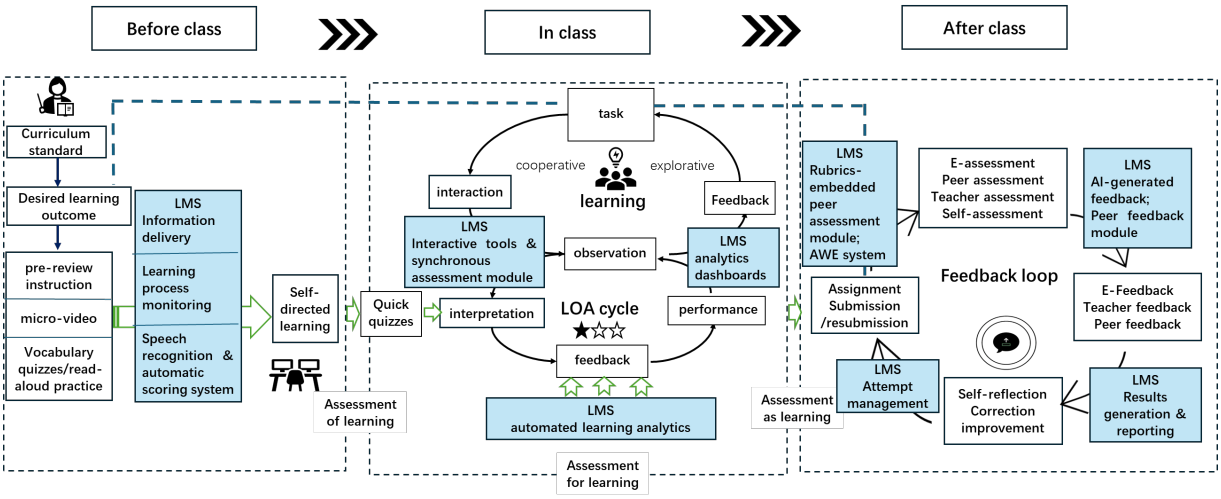


Figure 6. Technology-enhanced integration of LOA and the flipped classroom model

Nevertheless, as previously noted, the integration of technology into LOA remains a double-edged sword. Technological innovations, while substantially enhancing the validity, reliability, and efficiency of summative assessments and facilitating the effective capture of learning-process data, have also introduced challenges. Specifically, concerns have emerged about increased student disengagement in assessment activities and a rise in academic dishonesty (Cutillas et al., 2025), which persists despite the implementation of measures such as anti-farming mechanisms and online proctoring. As noted by Ndungu and Chepsergon (2024), the introduction of AI-powered content generation tools has fundamentally challenged academic integrity, making it easier for students to commit plagiarism and cheating, thereby hindering instructors' ability to accurately gauge students' genuine understanding and intellectual capabilities. This tension reflects the broader challenge of balancing the benefits of technology-mediated assessment with the need to safeguard academic integrity (Alessio & Messinger, 2021).

To explore this issue, UTAUT, which recognizes four determinants of user intention and behavior, namely effort expectancy, social influence, performance expectancy, and facilitating conditions, together with engagement theory, could be employed as a framework to diagnose academic dishonesty and the disengagement arising from the integration of technology (Pan & He, 2024). In this research, students' disengagement in technology-enhanced peer assessment may result from a perceived mismatch between the substantial effort involved in studying rubrics and delivering detailed peer feedback and the relatively low grading returns. This inference is supported by Lund Dean and Jolly (2012), who posit that disengagement may occur when perceived efforts outweigh the expected benefits. Conversely, research indicates that when technology enhances performance expectancy while reducing perceived effort, students may be more inclined to misuse it (Mahda et al., 2024). In this research, as students' online learning effort outside the classroom was also factored into final grades to incentivize students' engagement, students were inclined to cheat to get high grades by using automatic course completion tools. Moreover, limited teacher guidance on appropriate technology use, indicative of weak facilitating conditions, may have undermined students' engagement in technology-assisted assessment and feedback; when coupled with assessment tasks of misaligned difficulty, this may be more likely to render academic dishonesty.

Overall, the quantitative results and qualitative insights provide educators with clear evidence that the "revolutionary" improvement in student learning is primarily driven by a pedagogical shift toward a learning-oriented mentality rather than technology itself. They need to weigh the pedagogical affordances and ethical implications of technologies, particularly in how to leverage the affordances of technology to assist the LOA design while mitigating potential risks.

IMPLICATIONS FOR RESEARCH AND EDUCATION

Overall, through quantitative investigation and qualitative exploration, the research offers a thorough examination of the cause-and-effect relationship between LOA intervention and TEVT students' English language learning outcomes, while identifying the LOA strategies contributing to these outcomes, with a special focus on the mediating role of technology. It has significant implications for LOA practitioners, curriculum designers, and researchers.

IMPLICATIONS FOR LOA PRACTICE

The research demonstrated how rubric-driven peer assessment, action-oriented feedback loops, protocol-based self-assessment, and self-reflection could yield positive effects on language learning. These findings offer practical guidance for LOA practitioners to optimize the design of technology-mediated peer assessment, self-assessment, and feedback in language classes. Additionally, the technology-enhanced LOA integrated flipped classroom model, established through the qualitative exploration, provides a concrete and actionable framework for vocational college English curricular designers to improve assessment practice in a flipped English language learning context, particularly to address the challenges of large classes. To facilitate its implementation, the study underscores the importance of educational institutions providing training to TEVT teachers to enhance their capacity to design learning-oriented assessment tasks with the assistance of cutting-edge technology and to support a pedagogical shift from a grade-oriented to a learning-oriented mindset.

Moreover, the research is also meaningful to the administrative body and policymakers, who are responsible for developing curriculum standards to guide assessment and instruction practice in a centralized educational context. It demonstrated the potential of LOA to transform passive learners into active learners and yield positive learning outcomes, thereby informing policymakers' and educational stakeholders' decision-making in ongoing assessment innovation initiatives within an exam-oriented culture.

Implications for LOA Research

With solid evidence, the implementation of the LOA intervention was found to positively affect the language proficiency of TVET students in the flipped classroom environment in China. This efficacy was particularly pronounced in writing skills, evidenced by the statistically significant differences observed both within the experimental group (pre- vs. post-intervention) and between the experimental and control groups. These findings further validate that LOA principles, such as designing assessment tasks for learning, involving students in assessment activities, and using feedback for feedforward, are viable and effective in improving TVET students' learning achievements (Wakid et al., 2024). It contributes to the discussion of LOA effects on language proficiency in the specific context of the flipped English class in China's TEVT institutions.

This qualitative exploration delineates a holistic strategy for leveraging technology to realize the full potential of LOA, thereby addressing the research gap in comprehensive, systematic LOA implementations. This holistic strategy could inform future research on optimizing technology-enhanced LOA intervention. It demonstrates its relevance by contributing to the ongoing discussion regarding how to reconcile multiple roles of assessment in the educational system (William & Thompson, 2017). The research also initiated a discussion of the potential tensions inherent in using technology-mediated assessment, specifically the need to maximize its benefits while simultaneously establishing adequate safeguards for academic integrity (Alessio & Messinger, 2021). Additionally, UTAUT was employed as an analytical framework to identify the factors that may contribute to academic dishonesty and disengagement in the technology-integrated context. This theory-driven analysis extends the study's practical implications by offering diagnostic insights and providing a foundation for future research to develop and test targeted interventions.

Transferability of the LOA Model

Although the research was situated within the specific context of China's TEVT institutions, the LOA strategies adopted, including iterative feedback loops, rubric-driven peer assessment, and protocol-based self-assessment strategies, reflected the principles that have been widely validated by prior LOA studies beyond China. Moreover, the technology-enhanced LOA integrated flipped classroom model demonstrates theoretical coherence with established theories, including Self-Determination Theory, Active Learning Theory, and UTAUT, suggesting its relevance beyond immediate situational contexts. Thus, these elements are not context-dependent and can be generalized to broader contexts beyond China. The technology-enhanced implementation strategies are also applicable across diverse higher vocational education settings, particularly in contexts characterized by large class sizes and limited classroom instruction hours, provided that an integrated digital platform is available.

However, there are also context-dependent aspects of the model and strategies that require careful consideration. In China's TEVT system, assessment practices are guided and supported by the Teaching and Learning Advisory Committee, national curriculum standards, an LMS accessible to everyone, and faculty collaboration. These factors ensure sufficient class hours (45–90 per year) for in-class assessment and feedback, facilitate platform accessibility, accommodate students' technical literacy, and support formative assessment with immediate feedback. In the absence of policy and faculty support, designing bite-sized assessment tasks, creating protocols and rubrics, building item banks, and delivering quality teacher feedback would present substantial challenges.

Meanwhile, influenced by an exam-oriented culture, TEVT students in China may have been accustomed to frequent bite-sized assessments and may benefit from them, though they may also lead to stress and cognitive overload (Skulmowski & Xu, 2022; Vaessen et al., 2017). However, this potential negative effect could be amplified in other settings, potentially offsetting the pedagogical benefits of the practice. Given this, when transferring the LOA integrated flipped classroom model, teachers need to redesign assessment tasks to evaluate students' performance through alternative formats such

as presentations, project assignments, or written reports. On top of that, it should be noticed that English language learning is a non-linear and dynamic process (Ellis, 2008); thus, when transferring the LOA model to more linear subjects, the assessment cycle needs to be streamlined and more tightly sequenced.

LIMITATIONS

Despite the attainment of research objectives in this empirical study, limitations can be identified, suggesting directions for further research. As mentioned, acquiring English as a foreign language is a gradual and cumulative process that occurs over an extended period. Consequently, the 12-week intervention period of this study may have been insufficient to allow the full effects on students' language achievements to manifest. Meanwhile, as mentioned earlier, the majority of participants were female, which introduced a potential gender bias. The participants' backgrounds were relatively homogeneous. While such homogeneity helped reduce the influence of extraneous variables, it might also limit the diversity of the sample and consequently constrain the generalizability of the findings. Hence, it is suggested that future research investigate the impact of prolonged intervention on vocational college English learners with a high school background and more balanced gender distribution, a demographic that was not included as participants in the current study.

Despite a qualitative exploration of LOA strategies that are affecting TEVT students' language proficiency, the discussion is limited to the pedagogical scope; factors such as students' personal traits, previous learning experiences, attitudes, and perceptions were insufficiently addressed. Thus, further study should expand the insights of LOA strategies by including the perspectives of learners.

CONCLUSIONS

This research demonstrates that although integrating LOA into the flipped classroom did not yield a statistically significant improvement in overall English proficiency among vocational college students in Central China, it did yield significant gains in specific areas, particularly writing skills. These findings further validated that LOA practice is especially beneficial for the productive, complex skill that involves multiple mental processes and relies on feedback from audiences or readers to validate its quality (Imsa-ard, 2024). Meanwhile, they also revealed that learners with different traits and academic backgrounds may respond differently to LOA interventions. Further research with a larger, balanced-gender sample is needed to validate the effectiveness of LOA across different types of learners in TVET contexts with diverse backgrounds and a prolonged intervention period.

The study further revealed that the effectiveness of LOA on language proficiency was contingent upon several critical factors, including the design of assessment tasks, the provision of feedback, and the implementation of self-assessment and peer assessment. It demonstrated that an action-oriented feedback loop was conducive to promoting deep learning and improving writing skills. Self-assessment and self-reflection informed by feedback and aligned with desired learning outcomes were found to foster self-directedness and enhance learning outcomes. However, these benefits cannot be achieved without addressing the challenges inherent in the specific context of TVET in China. On one hand, institutions should provide targeted training for instructors on LOA literacy, AI-based assessment tasks development, and feedback systems to scale the practice. On the other hand, instructors should be empowered to calibrate learning objectives or desired learning outcomes in accordance with students' performance in ongoing assessments. Concurrently, given that students with low motivation and interest in English tend to get only limited benefits from LOA practices, further investigation is needed to examine the dynamics among the language learners' traits, the LOA intervention, and the instruction model.

It improves the reliability and validity of LOA tasks, facilitates the capture of impromptu learning data for formative assessment, and supports swift, evidence-based pedagogical adjustment. However, concerns over academic dishonesty and disengagement in assessment activities remain a persistent and evolving problem. Thus, the research recommended that further research focus on exploring how to maximize the benefits of technology in LOA practice while minimizing potential risks.

ACKNOWLEDGMENTS

This article is a component of the first author's doctoral research at the Faculty of Technical and Vocational Education (FTVE), Universiti Tun Hussein Onn Malaysia (UTHM). This study was made possible by the generous research funding received from the Advisory Committee on Foreign Language Teaching in Vocational and Technical Education, Ministry of Education, China, and the Shanghai Foreign Language Education Press (Contract Grant WYJZW-2023HB0016). Meanwhile, we would like to extend gratitude to Wuhan Technical University and all participants who not only volunteered to take part in this research but also generously offered the necessary assistance and support.

REFERENCES

Alessio, H. M., & Messinger, J. D. (2021, April). Faculty and student perceptions of academic integrity in technology-assisted learning and testing. *Frontiers in Education, 6*, 629220. Frontiers Media SA. https://doi.org/10.3389/feduc.2021.629220

Aliakbari, M., Barzan, P., & Sayyadi, M. (2025). Automated feedback vs. human feedback: A study on AI-driven language assessment. *AI and Tech in Behavioral and Social Science*, *3*(2), 113–126. https://doi.org/10.61838/kman.aitech.3.2.9

Al-Rashidi, A. H., Asif, M., Vanani, M. G., & Aberash, A. (2022). Learner-oriented assessment (LOA) practice: The comparative study of self-assessment, peer assessment, and teacher assessment on EFL learners' writing complicity, accuracy, and fluency (CAF), speaking CAF, and attitude. *Language Testing in Asia, 12*, Article 59. https://doi.org/10.1186/s40468-022-00209-x

Ambarcıoğlu, P. (2021). The use of effect size in veterinary medicine. *Veterinary Journal of Mehmet Akif Ersoy University*, *6*(1), 28–33."https://doi.org/10.24880/maeuvfd.823493

Awang, M. I. (2021). The digitalization of learning assessment. *Proceedings of the 4th International Conference on Multidisciplinary Research*, 4(1). https://doi.org/10.32672/pic-mr.v4i1.3731

Banerjee, H. L. (2021). Approaches to researching learning-oriented assessment in second and foreign language classrooms. In A. Gebril (Ed.), Learning-oriented language assessment (pp. 49–68). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003014102-5/approaches-researching-learning-oriented-assessment-second-foreign-language-classrooms-heidi-liu-banerjee

Banitalebi, Z., & Ghiasvand, F. (2023). The representation of Learning-Oriented Assessment (LOA) practice in EFL contexts: A case study of teacher questioning strategies. *Language Related Research*, *14*(3), 69–95.

Bartlett, J. (2019). *Introduction to sample size calculation using G*Power* (Version 1.1). OSF. https://files.osf.io/v1/resources/pcfvj/providers/osfstorage/5dcea84d5b97bd000e57aba0?action=download&direct=&version=1

Bickman, L., & Rog, D. J. (Eds.). (2009). *The SAGE handbook of applied social research methods* (2nd ed.). SAGE. https://doi.org/10.4135/9781483348858

Bin Mubayrik, H. F. (2020). New trends in formative-summative evaluations for adult education. *Sage Open*, *10*(3). https://doi.org/10.1177/2158244020941006

Brown, S. (2019). Foreword. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education: A handbook for academic practitioners* (pp. 1–3). Routledge. https://doi.org/10.4324/9780429506857

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, *159*, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66. https://doi.org/10.1080/14703290601081332

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z

Chang, M.-Y., & Hsu, L.-L. (2006). Qualitative research: An introduction to focus group methodology and its application. *Hu li za Zhi The Journal of Nursing*, *53*(2), 67–72. https://pubmed.ncbi.nlm.nih.gov/16602049/

Chapelle, C. A., & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511733116

Chen, Y. (2022). PRETCO: An English test for vocational and technical college students. *International Journal of TESOL Studies*, *4*(1), 213–227. https://doi.org/10.46451/ijts.2022.01.14

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment - Companion volume*. https://www.coe.int/en/web/common-european-framework-reference-languages

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson. https://eric.ed.gov/?id=ED594549

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage. https://collegepublishing.sagepub.com/products/designing-and-conducting-mixed-methods-research-3-241842

Cutillas, A. L., Balili, E. E., Rellin, E. C., Remo, M. H., Quindala, E. S., & Bacalso, F. D. (2025). Analyzing alternative academic assessments (AAAs) in universities with cheat-proofing behaviors and instructor-formulated mechanics in an AI-dependent setting. *Environment and Social Psychology*, *10*(5). https://doi.org/10.59429/esp.v10i5.3265

Dankel, S. J., & Loenneke, J. P. (2021). Effect sizes for paired data should use the change score variability rather than the pre-test variability. *Journal of Strength and Conditioning Research*, *35*(6), 1773–1778. https://doi.org/10.1519/JSC.0000000000002946

Derakhshan, A., & Ghiasvand, F. (2022). Demystifying Iranian EFL teachers' perceptions and practices of learning-oriented assessment (LOA): Challenges and prospects in focus. *Language Testing in Asia*, *12*, Article 55. https://doi.org/10.1186/s40468-022-00204-2

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal, 92*(2), 232–249. https://doi.org/10.1111/j.1540-4781.2008.00716.x

Er, H. K., & Farhady, H. (2023). The effects of learning-oriented assessment on academic writing. *International Journal of Assessment Tools in Education*, *10*(2), 197–217. https://doi.org/10.21449/ijate.1222410

Estaji, M., & Safari, F. (2023). Learning-oriented assessment and its effects on the perceptions and argumentative writing performance of impulsive vs. reflective learners. *Language Testing in Asia*, *13*, Article 31. https://doi.org/10.1186/s40468-023-00248-y

Fazel, I., & Ali, A. M. (2022). EAP teachers' knowledge and use of learning-oriented assessment: A cross-contextual study. *System*, *104*, 102685. https://doi.org/10.1016/j.system.2021.102685

Gao, Y. (2017). An empirical study on learning-oriented assessment. *Advances in Educational Technology and Psychology, 1*(2), 33–43. https://doi.org/10.23977/aetp.2017.12002

Gebril, A. (2018). Test preparation in the accountability era: Toward a learning-oriented approach. *TESOL Journal*, *9*(1), 4–16. https://doi.org/10.1002/tesj.302

Gebril, A. (2021). Learning-oriented assessment: Main issues and an overview. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 1–10). Routledge. https://www.routledge.com/Learning-Oriented-Language-Assessment-Putting-Theory-into-Practice/Gebril/p/book/9780367761141

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2017). *Research methods in applied settings: An integrated approach to design and analysis*. Routledge. https://doi.org/10.4324/9781315723082

González, C. J., López, E. O., & Morales, G. E. (2018). A cognitive tool to evaluate meaning formation of course contents: A learning-oriented assessment approach. *Proceedings of the 6th International Conference on Information and Education Technology* (pp. 52–55). Association for Computing Machinery. https://doi.org/10.1145/3178158.3178204

Gramcheva, L. (2020). The effective feedback challenge: How to begin developing self-regulated learners. In T. Byrom (Ed.), *Meeting the challenges of teaching and learning in 21st century higher education: Universal design* (pp. 164–181). Cambridge Scholars Publishing.

Greasley, P. (2008). *Quantitative data analysis using SPSS: An introduction for health and social sciences*. McGraw-Hill. https://ndl.ethernet.edu.et/bitstream/123456789/79496/2/Quantitative%20Data%20Analysis2.pdf

Greenspon, R. (2023). *Supporting teachers in designing learning-oriented assessment in a technology-enhanced learning environment* [Doctoral dissertation. Vytautas Magnus University]. https://www.lituanistika.lt/content/108932

Guo, L. M., Yang, X. M., & Zhang, Y. (2021). Research on design and practice of a data-driven five-dimensional support service framework for precision teaching. *e-Education Research*, *42*(4), 85–92.

Hamp-Lyons, L. (2017). Language assessment literacy for language learning-oriented assessment. *Papers in Language Testing and Assessment*, *6*(1), 88–110. https://doi.org/10.58379/LIXL1198

Hamzelou, Z., Alavi, S. M., & Nemati, M. (2022). Integration of learning-oriented assessment into writing instruction: A case of IELTS writing preparation class. *Teaching English as a Second Language Quarterly*, *41*(3), 57–87.

Höl, D., & Akman, E. (2023). A bibliometric analysis on "E-assessment in teaching English as a foreign language" publications in Web of Science (WoS). In Ö. Demirezen & M. Özkan (Eds.), *Handbook of research on perspectives in foreign language assessment* (pp. 329–355). IGI Global. https://doi.org/10.4018/978-1-6684-5660-6.ch016

Hussain, S., Abbasi, Q. H., Ansari, I. S., Qadir, J., & Imran, M. A. (2019, April). Online interactivity tools to support student-centered learning in large classes. *12th Annual University of Glasgow Learning and Teaching Conference, Glasgow, UK*. https://eprints.gla.ac.uk/179775/

Imsa-ard, P. (2024). *Learning-oriented assessment of L2 argumentative writing ability: Its impacts on EFL undergraduate students' writing ability and academic resilience*. Research Square.!https://doi.org/10.21203/rs.3.rs-4448642/v1

Jalilzadeh, K., & Coombe, C. (2023). Constraints in employing learning-oriented assessment in EFL classrooms: teachers' perceptions. *Language Testing in Asia*, *13*, Article 7. https://doi.org/10.1186/s40468-023-00226-4

Janke, S., Rudert, S. C., Petersen, Ä., Fritz, T. M., & Daumiller, M. (2021). Cheating in the wake of COVID-19: How dangerous is ad-hoc online testing for academic integrity?. *Computers and Education Open*, *2*, 100055. https://doi.org/10.1016/j.caeo.2021.100055

Jiang, L., & Qu, Y. (2024). Assessing core competencies of business English students in Chinese higher vocational colleges. *Scientific Reports, 14*, Article 29328. https://doi.org/10.1038/s41598-024-78898-y

Jones, N., Saville, N., & Salamoura, A. (2016). *Learning-oriented assessment: A systemic approach*. Cambridge University Press. https://www.cambridgeenglish.org/Images/639876-silt-volume-45.pdf

Junejo, S., Khatoon, S., & Jaleel, B. (2022). Active learning strategies and their effects on learning: a review. *Pakistan Languages and Humanities Review*, *6*(3), 742–755. https://doi.org/10.47205/plhr.2022(6-III)65

Karaman, P. (2024). Effects of using rubrics in self-assessment with instructor feedback on pre-service teachers' academic performance, self-regulated learning, and perceptions of self-assessment. *European Journal of Psychology of Education*, *39*(3), 2551–2574. https://doi.org/10.1007/s10212-024-00867-w

Kim, G. J. (2022). Korean secondary EFL learners' reading comprehension development through the student-generated reading comprehension test development: From the learning-oriented assessment approach. *Journal of Asia TEFL*, *19*(2), 414–430.

Ko, S., & Rossen, S. (2017). *Teaching online: A practical guide* (4th ed.). Routledge. https://doi.org/10.4324/9780203427354

Krebs, R., Rothstein, B., & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning*, *17*(2), 627–650. https://doi.org/10.1007/s11409-022-09302-1

Kunschak, C. (2020). CEFR, CLIL, LOA, and TBLT – Synergising goals, methods and assessment to optimise active student learning. In M. DeBoer & D. Leontjev (Eds.), *Assessment and learning in Content and Language Integrated Learning (CLIL) classrooms*. Springer. https://doi.org/10.1007/978-3-030-54128-6_4

Langove, S. A., & Khan, A. (2024). Automated grading and feedback systems: Reducing teacher workload and improving student performance. *Journal of Asian Development Studies, 13*(4), 123–130. https://doi.org/10.62345/jads.2024.13.4.16

Li, J., & Paulino, F. B. (2025). Determinants of oral English proficiency among Chinese vocational college students: A descriptive analysis. *Theory and Practice in Language Studies, 15*(6), 1796–1806. https://doi.org/10.17507/tpls.1506.08

Li, S. (2020). The impact of technology-mediated task-based language teaching on learners' second language speaking development. *Language Teaching Research, 24*(3), 291–310. https://doi.org/10.64152/10125/44739

Lund Dean, K., & Jolly, J. P. (2012). Student identity, disengagement, and learning. *Academy of Management Learning & Education*, *11*(2), 228–243. https://doi.org/10.5465/amle.2009.0081

Ma, J. (2025). Organizing the course content of China's English curriculum standards for vocational colleges – An attempt based on work process. *International Journal of Social Science and Education Research*, *8*(12), 188-195. https://doi.org/10.6918/IJOSSER.202512_8(12).0027

Ma, M. (2023). Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives. *Language Testing in Asia*, *13*, Article 33. https://doi.org/10.1186/s40468-023-00249-x

Mahda, A. A., Nik Fauzi, S. F., Dealwis, C., & Haris, N. (2024). Implementation of ChatGPT in moulding university students' writing. *Journal of Creative Practices in Language Learning and Teaching*, *12*(3), 82–96. https://ir.uitm.edu.my/id/eprint/115710

Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching and Learning Practice*, *13*(3), Article 7. https://doi.org/10.53761/1.13.3.7

Matyakhan, T., Chusanachoti, R., & Santos, J. A. L. (2024). "I think I speak without hesitation": Learning-oriented assessment to enhance English oral communication of Thai pre-service teachers. *rEFLections*, *31*(3), 966–989. https://doi.org/10.61508/refl.v31i3.275912

McBeth, M. (2015). Revising by numbers: Promoting student revision through accumulated points. *Journal of Response to Writing*, *1*(2), Article 3. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1061&context=journalrw

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40–49. https://www.jstor.org/stable/42923742

Meccawy, M., Alzahrani, A., Mattar, Z., Almohammadi, R., Alzahrani, S., Aljizani, G., & Meccawy, Z. (2023). Assessing EFL students' performance and self-efficacy using a game-based learning approach. *Education Sciences*, *13*(12), 1228. https://doi.org/10.3390/educsci13121228

Ministry of Education of the People's Republic of China. (2021, April 9). *English curriculum standard for higher vocational education*. http://www.moe.gov.cn/srcsite/A07/moe_737/s3876_qt/202104/t20210409_525482.html

Mumpuni, K. E., Hadi, S., & Suyanto, S. (2025). What interventions improve the effectiveness of technology-facilitated peer assessment? A meta-analysis. *Journal of Information Technology Education: Innovations in Practice, 24*, Article 16. https://doi.org/10.28945/5606

Navaie, L. A. (2018). The effect of learning-oriented assessment on learning pronunciation among Iranian EFL learners. *International Journal of Education and Literacy Studies*, *6*(2), 63–68. https://doi.org/10.7575/aiac.ijels.v.6n.2p.63

Ndungu, J. N., & Chepsergon, A. K. (2024). Dual role of AI in academic dishonesty and integrity management in the institutions of higher learning in Sub-Saharan Africa. *Journal of Research in Education and Technology, 2*(1), 69–79.

Nikou, S. A., & Economides, A. A. (2018). Mobile-based assessment: A literature review of publications in major refereed journals from 2009 to 2018. *Computers & Education*, *125*, 101–119. https://doi.org/10.1016/j.compedu.2018.06.006

Onasanya, S. A., & Ajamu, A. O. (2024). Formative and summative assessment using technology: A critical review. *Mimbar Ilmu*, *29*(1), 1–8. https://doi.org/10.23887/mi.v29i1.72042

Ozdem-Mertens, C., & Orhan, M. A. (2023, March 3). *Cheating detection with a cheating trap in online education*. PsyArXiv. https://doi.org/10.31234/osf.io/e5632

Pan, Y., & He, W. (2024). Research on the influencing factors of promoting flipped classroom teaching based on the integrated UTAUT model and learning engagement theory. *Scientific Reports*, *14*, Article 15201. https://doi.org/10.1038/s41598-024-66214-7

Powers, D., Schedl, M., & Papageorgiou, S. (2017). Facilitating the interpretation of English language proficiency scores: Combining scale anchoring and test score mapping methodologies. *Language Testing*, *34*(2), 175-195. https://doi.org/10.1177/0265532215623582

Purpura, J. E. (2021) A rationale for using a scenario-based assessment to measure competency-based, situated second and foreign language proficiency. In M. Masperi, C. Cervini, & Y. Bardière (Eds.*), Évaluation des acquisitions langagières: du formatif au certificatif, mediAzi*oni 32: A54-A96. https://mediazioni.sitlec.unibo.it/index.php/no-32-2021.html

Purpura, J. E., & Turner, C. E. (2018, July). *Using learning-oriented assessment in test development* [Invited workshop]. Language Testing Research Colloquium, Auckland, New Zealand.

Quyen, N. T. D., & Khairani, A. Z. (2017). Reviewing the challenges of implementing formative assessment in Asia: The need for a professional development program. *Journal of Social Science Studies*, *4*(1), 160–177. https://econpapers.repec.org/article/mthjsss88/v_3a4_3ay_3a2017_3ai_3a1_3ap_3a160-177.htm

Raghavjee, R., Govender, I., & Subramaniam, P. R. (2021). Learning analytics in higher education. In P. Ndayizigamiye, G. Barlow-Jones, R. Brink, S. Bvuma, R. Minty & S. Mhlongo (Eds.), *Perspectives on ICT4D and socio-economic growth opportunities in developing countries* (pp. 398-431). IGI Global. https://doi.org/10.4018/978-1-7998-2983-6.ch015

Roni, S. M., Merga, M. K., & Morris, J. E. (2019). *Conducting quantitative research in education*. Springer.

Ruan, T. T., Huang, P. Q., & Zeng, W. J. (2021). Exploration of intelligentized learning-oriented assessment model – A pedagogical trial with AI lecturer. *Educational Research and Experiment*, *2*, 69–75. https://d.wanfangdata.com.cn/periodical/jyyjysy202102011

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Vansteenkiste, M. (2023). Self-determination theory: Metatheory, methods, and meaning. In R. M. Ryan & M. Vansteenkiste (Eds.), *The Oxford handbook of self-determination theory* (pp. 3–30). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.013.2

Sadeghi, K. (Ed.). (2022). Technology in language assessment: An overview. *Technology-assisted language assessment in diverse contexts: Lessons from the transition to online testing during COVID-19* (pp. 3–15). Routledge. https://doi.org/10.4324/9781003221463

Saville, N. (2019). How can multilingualism be supported through language education in Europe? *Language Assessment Quarterly*, *16*, 464–471."https://doi.org/10.1080/15434303.2019.1676246

Saville, N. (2021). Learning-oriented assessment: Basic concepts and frameworks in using assessment to support language learning. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 13–33). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003014102-3/learning-oriented-assessment-nick-saville

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*, 171–196. https://doi.org/10.1007/s10648-021-09624-7

Susanto, A. (2025). Innovation in learning-oriented language assessment [Review of the book, *Innovation in learning-oriented language assessment,* by Sin Wang Chong and Hayo Reinders (Eds.) (2023)]. *Teaching English as a Second Language Electronic Journal, 28*(4). https://doi.org/10.55593/ej.28112r3

Tsushima, R. (2022). *Becoming active users of assessment in learning-oriented EAP classrooms: A mixed methods exploratory sequential design with wrap-up phase* [Doctoral dissertation, McGill University]. https://escholarship.mcgill.ca/concern/theses/1j92gd18p

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–272). De Gruyter Mouton. https://doi.org/10.1515/9781614513827-018

Vaessen, B. E., van den Beemt, A., van de Watering, G., van Meeuwen, L. W., Lemmens, L., & den Brok, P. J. (2017). Students' perception of frequent assessments and its relation to motivation and grades in a statistics course: A pilot study. *Assessment & Evaluation in Higher Education, 42*, 872–886. https://doi.org/10.1080/02602938.2016.1204532

Viengsang, R., & Wasanasomsithi, P. (2022). Effects of a learning-oriented reading assessment model on Thai undergraduate students' reading ability. *LEARN Journal: Language Education and Acquisition Research Network*, *15*(1), 709–747. https://so04.tci-thaijo.org/index.php/LEARN/article/view/256744

Voss, E. (2021). The role of technology in learning-oriented assessment. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 207–224). Routledge. https://doi.org/10.4324/9781003014102-14

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wakid, M., Sofyan, H., Widowati, A., & Zaida Ilma, A. (2024). Learning-oriented assessment: A systematic literature network analysis. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2366075

Wang, F. (2022). *The effectiveness of learner-oriented assessment on source use of reading-to-write tasks among secondary school EFL writers* [Doctoral dissertation, The University of Iowa].

Waskito, I., Wulansari, R. E., & Kyaw, Z. Y. (2021). The adventure of formative assessment with active feedback in vocational learning: The empirical effect for increasing students' achievement. *Journal of Technical Education and Training*, *14*(1), 54–62. https://doi.org/10.30880/jtet.2022.14.01.005

Wicking, P. (2017). Learning-oriented assessment in a testing-oriented culture. *The Language Teacher*, *41*, 9–13. https://doi.org/10.37546/JALTTLT41.4-2

William, D., & Thompson, M. (2017). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment* (2nd ed., pp. 53–82). Routledge. https://doi.org/10.4324/9781315086545-3

Wu, D. (2025, July). *Human–AI collaboration and innovative applications in educational scenarios* [Video]. Smart Education of China. https://core.teacher.vocational.smartedu.cn/p/course/vocational/v_1006810077178056704

Yagin, F. H., Pinar, A., & de Sousa Fernandes, M. S. (2024). Statistical effect sizes in sports science. *Journal of Exercise Science & Physical Activity Reviews*, *2*(1), 164–171. https://doi.org/10.5281/zenodo.12601138

Yamamoto, S. (2003). Can corrective feedback bring about substantial changes in the learner's interlanguage system? *Studies in Applied Linguistics and TESOL*, *3*(2), 1–9. https://doi.org/10.7916/salt.v3i2.1624

Yang, R. (2023). Is it possible to implement learning-oriented assessment principles in test preparation? Evidence from a high-stakes standardised EFL test in China. In S. W. Chong, & H. Reinders (Eds.), *Innovation in learning-oriented language assessment. New language learning and teaching environments*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-18950-0_3

Yang, S., & Berdine, G. (2021). Normality tests. *The Southwest Respiratory and Critical Care Chronicles, 9*(37), 87–90. https://doi.org/10.12746/swrccc.v9i37.805

Yang, X. (2020). The cultivation of Chinese learners' English writing competency through learning-oriented assessment. *Theory and Practice in Language Studies*, *10*(10), 1227–1233. https://doi.org/10.17507/tpls.1010.06

Yang, Y., & Mohd Ahyan, N. A. (2025). Optimizing project-based flipped learning in Chinese TVET English: Key elements, student satisfaction, and admission pathway effects. *International Journal of Academic Research in Progressive Education and Development*, *14*(1), 1724–1745. https://doi.org/10.6007/IJARPED/v14-i1/24707

Yoon, S. Y., & Bhat, S. (2018). A comparison of grammatical proficiency measures in the automated assessment of spontaneous speech. *Speech Communication*, *100*, 44–55. https://doi.org/10.1016/j.specom.2018.04.006

Yoqubovna, B. N. (2025). The role of the NOW Corpus in developing students' language skills. *EduVision: Journal of Innovations in Pedagogy and Educational Advancements, 1*(3), 941–946. https://brightmindpublishing.com/index.php/ev/article/view/380

Yu, R. R. (2014). Corrective feedback in English class and interlanguage development. *Overseas English, 2*(X), 31–33.

Yuan, C., & Moran, C. M. (2018). Flipped classroom in China: Design, practice, and implications. In H. A. Spires (Ed.), *Digital transformation and innovation in Chinese education* (pp. 119–135). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-5225-2924-8.ch007

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning: A critical review of literature. *Educational assessment, evaluation and accountability*, *30*, 211–250. https://doi.org/10.1007/s11092-018-9281-9

Zhao, C. G., & Qi, Q. (2023). Implementing learning-oriented assessment (LOA) among limited-proficiency EFL students: Challenges, strategies, and students' reactions. *TESOL Quarterly, 57*(2), 566–594. https://doi.org/10.1002/tesq.3167

Zhao, Y., Sulaiman, N. A., & Wahi, W. (2024). Longitudinal investigation: Impact of production-oriented approach on Chinese university students' English writing and speaking proficiency. *World Journal of English Language*, *14*(4), 92–100. https://doi.org/10.5430/wjel.v14n4p92

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese undergraduates. *Assessing Writing, 37,* 13-24. https://doi.org/10.1016/j.asw.2018.11.001

Zhou, M. (2023). Significance of assessment in learning: The role of educational assessment tools. *Science Insights Education Frontiers*, *18*(2), 2881–2883. https://doi.org/10.15354/sief.23.co215

Zúñiga, J. L. B. (2024). *Examining assistance in a scenario-based test of L2 argumentative speaking ability: A learning-oriented approach for assessing complex competencies in language education* [Doctoral dissertation, Columbia University].

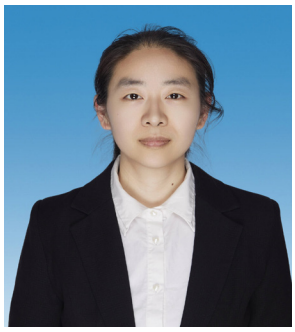
Zuo, X.-y. (2016). Motivation in a flipped classroom, a case study of teaching oral English in a vocational college in mainland China. *Sino-US English Teaching*, *13*(6), 460–467.

AUTHORS

Jing Zha is an associate professor of English Education at the School of Foreign Languages and International Trade, Wuhan Technical University, China. She is currently pursuing her doctoral studies in Technical and Vocational Education at the University Tun Hussein Onn, Malaysia (UTHM). Her research focuses primarily on English education, technology-mediated learning, learning assessment, applied linguistics, and teacher professional development, with a special interest in flipped classroom instruction, content- and language-integrated learning, and learning-oriented assessment. She has published approximately 40 articles in Chinese journals and completed 5 research projects in these areas, demonstrating her commitment to advancing English pedagogy and maximizing the affordances of technology for language acquisition.

Azita binti Ali obtained a bachelor's degree in data and communication from University Technology MARA (UiTM) and subsequently earned a Doctor of Philosophy (PhD) in Technical and Vocational Education from University Technology Malaysia (UTM). In December 2006, she served as a tutor at the Faculty of Technical and Vocational Education (FPTV). After completing her PhD studies in 2012, she was appointed as a lecturer at FPTV, UTHM. She has served as Head of the Department of Vocational Education since October 2023. With 18 years of experience as a lecturer, her main research interests focus on the development of teaching and learning materials, particularly in Creative Multimedia and Technical and Vocational Education. She has also published numerous scholarly articles in both national and international journals in her field of expertise.

Yan Tan is a lecturer at Wuhan Technical University. She holds a master's degree in education from Hubei Normal University and currently teaches public English courses at Wuhan Technical University. She has dedicated nearly two decades to English vocational education and research within TVET institutions. Her research focuses on English pedagogy, the flipped English classroom, curriculum design, technology-assisted learning, and assessment, with particular attention to learning-oriented assessment research and practice based on smart learning management systems, such as integrating learning-oriented assessment into flipped classroom instruction in vocational English courses. She has published multiple papers in Chinese journals in related fields and has participated in several research projects.



Yuting Wang is a lecturer at the Department of Public English Teaching, Wuhan Technical University. Her primary research interests lie in English instruction strategies and technology-enhanced assessment, with a particular focus on second language acquisition, technology-assisted learning and instruction, and learning-oriented assessment. With 20 years of experience in English teaching, she has an impressive academic record, including multiple papers published in Chinese academic journals. She has participated in several key educational research projects in these fields. Her research findings offer valuable insights into optimizing pedagogical practices for public English courses within TVET contexts.