INTERNATIONAL JOURNAL OF TESOL & EDUCATION

ISSN 2768-4563

ICTE Press
International Association of TESOL & Education
5243 Birch Falls LN, Sugar Land, Texas, USA, 77479
https://i-jte.org | https://i-cte.org/press

International Journal of TESOL & Education (ijte)

ISSN: 2768-4563
Volume 5, Issue 2, 2025

Doi: https://doi.org/10.54855/ijte.2552
Orcid: https://orcid.org/0000-0002-3233-457X

Indexed
Crossref
Google Scholar
LOCKSS
CLOCKSS
PKP PN
Library of Congress, U.S. ISSN Center
ROAD
J-Gate
EuroPub

Publication frequency
International Journal of TESOL & Education will publish 4 issues per year in February, May, August, and November. The IJTE will publish manuscripts as soon as they are accepted.

Publisher: ICTE Press
Address: International Association of TESOL & Education, 5243 Birch Falls Ln, Sugar Land, Texas, USA, 77479

Principal Contact
Associate Professor Dr. Pham Vu Phi Ho
Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam
Email: phamvuphiho@iuh.edu.vn; ijte.editorial@gmail.com

International Journal of TESOL & Education

Vol. 5, No. 2 (2025): TESOL & Education

Editorial Board

International Journal of TESOL & Education
Volume 5, Issue 2, 2025

Reviewers

Dr. Kurt Candilas, Lourdes College, Cagayan de Oro City, Philippines

Nguyen Huy cuong, National Taiwan University of Science and Technology, Taiwan

Dr. SettingsTran Thi Minh Phuong, HCMC University of Social Sciences and Humanities, Vietnam National University, Vietnam

Tran Quang Hai, Hoa Sen University, Vietnam

Nguyen Quynh Thy, HCMC University of Social Sciences and Humanities, Vietnam National University, Vietnam

Tran Thi Ngoc Linh, Faculty of Foreign Languages, Van Lang University, Vietnam

Trang Tran, Hoa Sen University, Vietnam

Tran Vu Diem Thuy, Ho Chi Minh City University of Foreign Languages - Information Technology (HUFLIT), Vietnam

Le Duc Hanh, Hanoi University of Industry, Hanoi, Vietnam

Tran Thi Thanh Mai, Faculty of Foreign Languages, Van Lang University, Vietnam

Ngo Thi Cam Thuy, Faculty of Foreign Languages, Van Lang University, Vietnam

Pham Thanh Su, Institute of Foreign Languages, University of Economics and Law, Vietnam

Cao Thi Xuan Tu, Faculty of Foreign Languages, Van Lang University, Vietnam

Nguyen Huu Hoang, Academy of Journalism and Communication, Vietnam

Phan Thi Kim Thao, HCM City University of Industry and Trade, Vietnam

Pham Van Thoa, Dong Nai University, Dong Nai Province, Vietnam

Dr. Vuong Thi Hai Yen, Thu Do University, Ha Noi, Vietnam

Nushrat Mahjabeen, University lecturer, Bangladesh

Dr. Tran Cao Boi Ngoc, HCMC University of Social Sciences and Humanities, Vietnam National University, Vietnam

Bui Thi Ngoc Oanh, Nha Trang University, Nha Trang City, Vietnam

Table of Content

Vol. 5 No. 2 (2025): TESOL & Education
Doi: https://doi.org/10.54855/ijte.2552
Orcid: https://orcid.org/0000-0002-3233-457X

Research Articles ***Pages***

Literature Review

A Note from the Editor-in-Chief

Dear beloved TESOLers & Educators,

With great pleasure, we announce the successful completion and publication of Vol. 5, No. 2, 2025 of the International Journal of TESOL & Education (IJTE). This issue includes eight research papers that cover crucial topics in language education, teacher preparation, and the changing incorporation of artificial intelligence technologies in learning and evaluation. Studies on this area demonstrate how dynamic the TESOL and education sectors are. They study the intersection of artificial intelligence applications, cooperative learning approaches, and competency-based teacher preparation, providing an informed analysis of the opportunities and problems in modern education.

In the rapidly changing field of language education, the articles in this issue have looked into many aspects of artificial intelligence integration, cooperative learning methodologies, and teacher professional development to meet modern educational needs. Quach and Nguyen (2025) discovered significant areas for improvement in General English Teaching Programs in Vietnam by investigating lecturers' perspectives on quality assurance (QA). Dinh (2025) used Dynamic Systems Theory to analyze the language development trajectories of Vietnamese EFL learners, revealing complicated, nonlinear processes. Yang (2025) studied the usefulness of the Fish-Skeleton Vocabulary Learning Diagram (FSVLD) combined with crossword puzzles for vocabulary retention among Taiwanese EFL students. Phan and Luong (2025) investigated the challenges that Legal English students faced while using AI approaches, focusing on ethical concerns and the danger of overreliance. Pham and Tran (2025) assessed students' perceptions of AI language models as virtual writing assistants in terms of grammatical help and idea development, taking into account concerns about accuracy and plagiarism. While Nguyen et al. (2025) explored factors influencing the efficacy of teacher training on competency-based assessment in Northern Vietnam, Nguyen et al. (2025) conducted a systematic study of the benefits and limitations of artificial intelligence translation tools in higher education. Pham (2021) demonstrated how students' writing fluency improved under cooperative writing structures. These publications, which address technology innovations, pedagogical techniques, and evaluation reforms, collectively contribute to the advancement of language education research.

In this context, Quach and Nguyen (2025) explored how lecturers at two Vietnamese universities perceived Quality Assurance (QA) in General English Teaching Programs (GETs). Using an explanatory mixed-methods approach and 76 lecturers, The data came from document analysis, surveys, and interviews. The results reflect varying levels of QA awareness; one university had superior QA performance. The report contains helpful information for improving quality assurance systems in Vietnamese higher education.

Dinh (2025) replicated Larsen-Freeman's (2006) study to look into how correctness, complexity, and fluency evolved in the written works of Vietnamese upper-intermediate EFL students. Four participants' narratives were analyzed for language development over a three-month period using a time-series methodology. The data revealed shifting trajectories that reflect DST's view of language as a nonlinear, interconnected system shaped by contextual factors. This study demonstrates the applicability of DST concepts in EFL settings and emphasizes the need of context-sensitive language education.

Yang (2025) at Fooyin University investigated the effectiveness of integrating crossword puzzles and the Fish-Skeleton Vocabulary Learning Diagram (FSVLD) in teaching EFL vocabulary in Taiwan. The quasi-experimental design research of 71 junior college students found that the intervention group had much superior vocabulary acquisition and retention than the control group receiving standard instruction. The findings demonstrate how well visual aides and puzzles can boost long-term memory and word acquisition.

Investigating issues experienced by legal English students utilizing AI tools for legal writing, Phan and Luong (2025) at Ho Chi Minh City University of Law looked at With 42 students, the study found using a survey methodology overreliance on artificial intelligence, plagiarism hazards, and diminished originality as main concerns. Results highlight the importance of pedagogical direction in guaranteeing responsible and efficient AI application in legal writing education.

At Van Lang University, Pham and Tran (2025) asked students about their impressions of artificial intelligence language models as virtual writing helpers in English writing classes. Under a mixed-methods approach, they polled 147 students—mostly English majors. Results showed good opinions on how artificial intelligence might improve grammar, vocabulary, and idea development but raised questions about accuracy, overreliance, and plagiarism.

At University of Science and Technology - The University of Da Nang, Nguyen (2025) investigated language and knowledge content gaps in IELTS academic writing. Descriptive and qualitative techniques were applied in analyzing 202 writings from 101 engineering freshmen. Emphasizing the requirement of teaching academic vocabulary, argumentation, and cross-cultural examples in writing, results exposed substantial grammar, vocabulary, and topic deficiencies.

At the University of Social Sciences & Humanities, Vietnam National University Ho Chi Minh City, Vietnam, Nguyen et al. (2025) conducted a comprehensive review at Vietnam National University Ho Chi Minh City to investigate the benefits and downsides of AI translation tools in tertiary-level translating courses. Combining 20 peer-reviewed studies (2014–2024), they found seven drawbacks including overreliance, accuracy problems, and ethical concerns alongside eight positives including better efficiency, vocabulary, and motivation.

Nguyen, Le, and Bui (2025) looked at elements affecting the results of a seven northern Vietnamese province competency-based evaluation teacher training program. Using regression analysis, they asked 1,422 teachers about four main elements: online support following training, training organization, IT infrastructure, and training materials. The study determined that training efficacy was most influenced by online support.

Emphasizing the transforming power of combining artificial intelligence tools, cooperative techniques, and competency-based assessment methodologies, the collection of papers offers important insights into the dynamic and changing terrain of language teaching. Although technological developments—especially AI tools for writing, translation, and learning support—offer significant benefits in efficiency, vocabulary acquisition, and learner motivation across these studies, they also present challenges related to overreliance, accuracy, ethical concerns, and contextual adaptation. Especially Nguyen et al. (2025) and Phan and Luong (2025) stress the need of striking a balance between artificial intelligence support with critical thinking and human judgment to guarantee appropriate use. Likewise, Pham and Tran (2025) and Yang (2025) show that although artificial intelligence and structured visual frameworks (such as FSVLD) could increase engagement and learning gains, pedagogical support and training remain very vital for optimizing efficacy. Particularly for diverse student groups, the studies on teacher preparation—especially Nguyen et al. (2025)—showcase the vital need of post-training support, customized content, and contextual awareness. Pham (2021) supports even more the need of organized group writing in improving fluency.

These research taken together show how important artificial intelligence technology, cooperative systems, and customized teacher preparation are in developing language instruction. The findings warn against overreliance and stress the need of critical thinking, contextual adaptation, and ethics even if artificial intelligence provides efficiency and improved learning opportunities. Equitable and successful professional development depends on teacher training programs, especially those combining post-training support and culturally sensitive approaches. These efforts taken together highlight a balanced, inclusive, and learner-centered approach in tackling the difficulties of twenty-first-century education.

We especially thank all authors and their affiliated institutions for their invaluable contributions; among them are scholars from Hanoi Pedagogical University 2, Van Lang University, Fooyin University, University of Science and Technology – The University of Da Nang, University of Social Sciences & Humanities – Vietnam National University Ho Chi Minh City, and Ho Chi Minh City University of Law. We also truly appreciate the time, knowledge, and dedication of our committed peer reviewers as well as the whole IJTE editorial team to help to preserve the integrity and quality of this publication.

The International Journal of TESOL & Education is currently accepting submissions for Vol. 5, No. 3, 2025. We are seeking original research papers, literature reviews, and practical reports that cover all aspects of TESOL and education. Authors who demonstrate enthusiasm are encouraged to submit their works directly. Our website, https://i-jte.org, contains the author guidelines and submission instructions.

We look forward to receiving your contributions and continuing to serve as a platform for academic discourse and innovation in language education.

Thank you for your continued support and dedication to the International Journal of TESOL & Education. We look forward to bringing you more exceptional content in the future.

Thanks be to God for everything!

Warm regards,

Associate Professor Dr. Pham Vu Phi Ho
Editor-in-chief
International Journal of TESOL & Education

Lecturers' Perceptions of the Quality Assurance System for General English Teaching Programs in Vietnamese Higher Education

Quach Thi To Nu[1,2*], Nguyen Loc[1]

[1]HCM City University of Social Sciences and Humanities - Vietnam National University, Vietnam
[2]Industrial University of Ho Chi Minh City, Vietnam
*Corresponding author's email: tonuquachthi@gmail.com
https://orcid.org/0000-0002-9297-2733
https://doi.org/10.54855/ijte.25521

Received: 19/01/2025 Revision: 20/04/2025 Accepted: 23/04/2025 Online: 23/04/2025

ABSTRACT

Understanding lecturers' perceptions of Quality Assurance (QA) is vital for improving English language programs. Despite growing attention to QA in higher education, limited research explores how lecturers view its implementation. This study addresses that gap by examining QA practices in General English Teaching programs at two universities in Ho Chi Minh City, Vietnam (HEI1 and HEI2). Using an explanatory mixed-methods approach, data were collected through questionnaires, interviews, and document analysis. Findings showed that lecturers from both institutions acknowledged QA efforts in curriculum design, teaching, assessment, staff development, and facilities. However, HEI2 consistently scored higher across most criteria. Qualitative insights highlighted HEI2's strengths in placement testing, workload design, staff training, and technology use. The study suggests HEI1 could benefit from benchmarking these practices. By involving lecturers, the research contributes to understanding QA implementation and offers practical recommendations for enhancing English programs in Vietnamese higher education.

Keywords: quality assurance, General English Teaching Program, higher education, lecturers' perception

Introduction

Quality assurance (QA) in higher education is vital for ensuring the effectiveness and relevance of academic programs. Harvey and Green (1993) describe quality in education through lenses such as fitness for purpose, value for money, and transformative potential. Robust QA mechanisms like program accreditation and regular evaluations help institutions meet defined standards (Martin & Stella, 2007). Continuous feedback from stakeholders—students, faculty, and employers—aligns program outcomes with societal and industry needs (Tam, 2001). Technology integration in QA processes enhances efficiency in monitoring and evaluation (Coates, 2005). Additionally, context-specific QA frameworks address diverse institutional and regional challenges (Biggs, 2001). A holistic approach to QA prioritizes compliance with standards and enhances the students' learning experience.

CITATION | Quach, T. T. N., & Nguyen, L. (2025). Lecturers' Perceptions of the Quality Assurance System for General English Teaching Programs in Vietnamese Higher Education. *International Journal of TESOL & Education*, *5*(2), 1-28. DOI: https://doi.org/10.54855/ijte.25521

In recent years, Quality Assurance (QA) has emerged as a cornerstone in the development and evaluation of academic programs across higher education institutions in Vietnam. QA frameworks are designed to ensure that educational programs meet institutional goals, stakeholder expectations, and international standards. In the context of General English Teaching Programs (GETPs), effective QA implementation is vital for maintaining program quality and achieving desired learning outcomes.

Despite the formal integration of QA systems into academic structures in Vietnam, many university lecturers remain unfamiliar with their roles and responsibilities in the QA process. This lack of understanding can hinder the successful implementation and effectiveness of QA initiatives. As QA principles such as accountability, continuous improvement, and stakeholder involvement are essential (Harvey & Green, 1993), the perceptions and engagement of lecturers who are key stakeholders in the teaching and learning process must be thoroughly understood.

Existing literature highlights several mechanisms that support effective QA, such as accreditation processes, program reviews, and benchmarking against international standards like the CEFR (Council of Europe, 2020). However, challenges such as limited resources (Owlia & Aspinwall, 1996), diverse stakeholders' needs (Tam, 2001), and resistance to change (Newton, 2000) often impede implementation. To overcome these barriers, institutions are advised to promote stakeholder involvement, offer professional development opportunities, and apply technological tools to improve QA efficiency (Kohoutek, 2009; Pereira et al., 2021).

Given this context, examining lecturers' perceptions of QA systems for General English Teaching Programs (GETPs) in Vietnam is both timely and necessary. The purpose of the study entitled "Lecturers' Perceptions of the Quality Assurance System for General English Teaching Programs in Vietnamese Higher Education" is to explore how lecturers perceive the implementation of QA practices, including their understanding, challenges encountered, and suggestions for improvement. This study may reveal lecturers' awareness of QA, obstacles they face, and opportunities to enhance QA practices in reality. Ultimately, understanding lecturers' perceptions will contribute to strengthening QA efforts, leading to better program outcomes, increased student satisfaction, and enhanced institutional reputation and graduate employability (Martin & Stella, 2007; Schindler et al., 2015).

Literature Review

General English Teaching Programs in higher education

To equip tertiary students with essential English language skills that are crucial for academic success and global communication, many HEIs in Vietnam provide General English programs. These programs often cater for non-English major students and are designed to develop their competencies in reading, writing, listening, and speaking skills. This section reviews recent studies to highlight their relevance and effectiveness in higher education.

GETPs are characterized by a curriculum that emphasizes communicative competence (Richards, 2006). These programs often adopt a skill-based approach, integrating real-life communication tasks to enhance learners' language proficiency (Brown, 2014). In many contexts, GETPs serve as foundational courses, preparing students for discipline-specific English or English for Specific Purposes popularly known as ESP programs (Hutchinson & Waters, 1987). In Vietnam, English is taught as a compulsory subject in the tertiary teaching programs other than English Language (Hoang, 2010). For instance, students might accumulate from 2 to 20 credits among more than 140 credits in their undergraduate curriculum. As stated

in the Higher education law and Vietnamese Language Proficiency Framework, B1 is the exit level for students as they wish to complete the bachelor degree.

Literature has discovered several challenges for GETPs. First, large class sizes and diverse proficiency levels make it difficult for instructors to address individual learning needs (Chen & Goh, 2011). Second, a lack of motivation among students, particularly those who do not perceive English as directly relevant to their fields of study, poses a significant challenge (Dörnyei, 2001). Third, limited essential resources and outdated teaching materials often negatively affect the quality of instruction (Gao, 2013). The context of GETPs in Vietnam also faces similar problems (Hoang, 2010).

Despite these difficulties, GETPs have found their ways to thrive in technology era. The integration of technology and blended learning, such as Learning Management Systems (LMS) and mobile applications, facilitates personalized learning experiences (Sun & Yang, 2021). Flipped classroom models, where students interact with online instructional contents before class, have also demonstrably improved learner engagement and outcomes (Wang, 2017). This technological shift complements pedagogical advancements like task-based language teaching (TBLT), which emphasizes authentic tasks and real-world applications. Research indicates that TBLT boosts students' communicative competence and critical thinking skills (Ellis, 2003).

Empirical evidence supports the effectiveness of well-designed GETPs in improving language proficiency. A meta-analysis by Zhang and Yin (2019) revealed that programs incorporating interactive activities and continuous assessment outperform traditional lecture-based models. Furthermore, aligning course objectives with the Common European Framework of Reference for Languages strengthens program coherence and assessment validity (Council of Europe, 2020). To ensure quality, Quach and Nguyen (2024) suggest benchmarking English training programs in Vietnam against established standards like NEAS.

All in all, while GETPs face challenges such as resource constraints and learner diversity, they play a crucial role in higher education by equipping students with essential language skills. The incorporation of technology innovation and task-based approaches, have demonstrably improved program effectiveness. Ultimately, sustainable implementation strategies are key to maximizing the long-term impact of GETPs on students' academic and professional success.

Quality Assurance in General English Teaching Programs

Quality assurance in GETP's objectives

Defining clear desired program objectives is essential to guarantee the training quality of an institution. Anderson and Krathwohl (2001) found that well-articulated objectives provide a roadmap for curriculum design, teaching contents, teaching methodologies, and assessment methods. They are claimed to ensure alignment between program goals, student needs, and institutional benchmarks, promoting coherence and consistency in delivery (Biggs & Tang, 2011). These objectives also serve as benchmarks for evaluating program effectiveness and pinpointing areas for enhancement (Tyler, 1949). Critically, when aligned with industry needs, they boost graduate employability by equipping students with relevant skills and knowledge (Barrie, 2006). Harden (2002) emphasized that objectives not only guide instructors in shaping instructional strategies but also clarify expectations for students, fostering active engagement and self-directed learning. The importance of SMART (Specific, Measurable, Achievable, Relevant, and Time-bound) objectives in upholding quality assurance in educational programs has been widely recognized by Doran (1981) and O'Neill (2020). In essence, establishing robust program objectives is a cornerstone of achieving desired learning outcomes and maintaining high quality tertiary training.

Quality assurance in GETP's learning outcomes

Learning outcomes play a pivotal role in ensuring the quality of training programs at the tertiary level as they provide a clear articulation of what students are expected to know, do, and value upon completing a program. This, therefore, serves as a roadmap for curriculum design, instructional methods, and assessment practices (Biggs & Tang, 2011). By defining specific and measurable outcomes, institutions can align teaching strategies with desired competencies, ensuring that students acquire relevant skills and knowledge (Anderson & Krathwohl, 2001). Learning outcomes also facilitate transparency, enabling stakeholders including students, educators, and employers to understand the purpose and scope of the program (Harden, 2002). Learning outcomes provide a foundation for evaluating program effectiveness and identifying areas for improvement (Kennedy, Hyland, & Ryan, 2009). They also encourage active learning and self-regulation by helping students focus on key tasks and monitor their progress (Biggs, 2003). Moreover, when aligned with industry standards, these outcomes boost graduate employability by connecting academic training with real-world professional needs (Barrie, 2006). As Dang and Pham (2024) pointed out, it is essential to use precise verbs in defining learning outcomes to avoid ambiguity and ensure effective teaching, learning and assessment. They also caution against overloading individual outcomes, stressing the importance of balance for a manageable and productive learning experiences. This focus on clear, relevant, and achievable outcomes, which suggests that modern curriculum design often employs "Backward Design" approach, starting with desired learning outcomes and planning the curriculum as well as the instruction afterwards (Wiggins & McTighe, 2005). In short, well-crafted learning outcomes are absolutely essential for maintaining and improving the quality of tertiary training programs.

Quality assurance in GETP's entry requirements

Entry requirements play a vital role in ensuring quality tertiary training by laying the groundwork for academic success and program coherence. Clear and appropriate admission criteria, as Yorke and Longden (2004) suggest, ensure incoming students possess the foundational knowledge, skills, and competencies necessary to effectively engage with the curriculum. This aligns with Kuh et al.'s (2006) observation that well-defined entry requirements improve student retention and performance by minimizing the mismatch between student preparedness and program demands. Furthermore, entry requirements contribute to the overall quality assurance framework by helping maintain academic standards and institutional reputation (Harvey & Green, 1993). Galloway (2009) highlights the importance of aligning entry requirements with program learning objectives to support student progression and completion rates by creating a better academic fit. Importantly, these requirements often reflect broader institutional goals, such as promoting diversity or addressing labor market needs, ensuring program relevance and inclusivity (Smith & Naylor, 2001). By carefully designing entry criteria, higher education institutions can strike a balance between accessibility and quality, building a strong foundation for student success and program success.

Quality assurance in GETP's structure and content

The structure and content of a program are fundamental in defining the quality of a program. First, a well-organized program structure provides a logical sequence of courses, ensuring foundational knowledge and skills are developed in accordance with the stated learning outcomes and programme outcomes. (Biggs & Tang, 2011). This "scaffolding" approach supports effective learning and facilitates the success of program outcomes. Furthermore, relevant and coherent program content is crucial for preparing students to meet academic, professional, and societal demands (Barnett, 2000). Alvesson and Sandberg (2013) emphasized

the need for program content to be aligned with the requirements of labor markets and global trends, which is served to equip graduates' employability and adaptability. A clear structure and relevant content also boost student engagement, as students perceive the program as purposeful and directly related to their goals (Merrill, 2002). Regular reviews and updates of program structure and content are required to maintain alignment with technological advancements and evolving knowledge domains (Fry et al., 2008), which ensures the program to be innovative, competitive, and capable of addressing current and future challenges. Therefore, a well-structured and contextually relevant program significantly contributes to the quality and effectiveness of tertiary education.

Quality assurance in GETP's learning volume

The learning volume requirement is a critical factor in ensuring quality tertiary training, as it defines the necessary workload to gain desired learning outcomes. This encompasses the total time students are expected to dedicate to learning activities, including lectures, self-study, and assessments, promoting a balanced and structured educational approach (Biggs & Tang, 2011). Well-defined learning volume requirements foster consistency across programs and institutions, facilitating comparability and transferability of qualifications (Adam, 2004). They also assist institutions in aligning their curricula with national or international credit frameworks, such as the European Credit Transfer and Accumulation System (ECTS), which standardizes workload and enhances program quality (European Commission, 2015). Research indicates that an appropriate workload prevents student from burnout and improves academic performance by providing sufficient time for reflection and content mastery (Kember, 2004). Learning volume requirements also contribute to transparency and accountability, providing students and stakeholders with clear expectations about the program demands (O'Neill, 2020). Consequently, designing and implementing learning volume requirements tailored to students' capabilities and program objectives is essential for sustaining the quality and effectiveness of tertiary education. In credit based curriculums, learning volume is clearly stated in Vietnam Qualifications Framework (Government of Vietnam, 2016) as a QA guideline for curriculum designer to comply to.

Quality assurance in GETP's teaching and assessment methods

Teaching methods and assessment strategies have been proved to be integral to ensuring the quality of training programs at the tertiary level. Biggs & Tang (2011) affirm that effective teaching methods foster student engagement, facilitate active learning, and support the achievement of desired learning outcomes. In particular, pedagogical approaches such as problem-based learning, outcome-based learning, collaborative projects, and experiential learning have been shown to enhance critical thinking and practical skills (Prince, 2004; Driscoll & Wood, 2007; Ho & Ha, 2025). In addition, self-paced learning has been proved to be especially appropriate to tertiary learners (Johnson et al., 2020; Balabag & Cadilas, 2024). Equally important are assessment practices, which serve as tools for measuring learning progress, providing feedback, and ensuring accountability (Boud & Falchikov, 2007). Aligning assessments with learning outcomes and teaching methods through constructive alignment ensures coherence and improves the success of learning and teaching (Biggs, 2003). Formative assessments, in particular, play a critical role in supporting student development by offering timely feedback and opportunities for improvement (Sadler, 1989; Brookhart, 2023). Meanwhile, summative assessments provide a basis for evaluating the overall effectiveness of a program and its ability to meet academic and professional standards (Brown & Knight, 1994; Gu & Lam, 2023). Furthermore, diverse and inclusive assessment methods cater for different learning styles and promote equity (Gibbs & Simpson, 2004; Carless, 2023). Dang and Tong

(2024) identified key instruction-giving techniques for Vietnamese EFL classrooms, including attention-grabbing, repetition, using the mother tongue, demonstrations, and checking understanding. Consequently, the thoughtful integration of innovative teaching methods like online resources accessed through personal technological devices (Nguyen, 2024) and robust assessment practices (Yastıbaş & Takkaç, 2018) is essential for maintaining high-quality tertiary education.

Quality assurance in GETP's academic staff

Academic staff are essential for ensuring the quality of the training programs because their qualifications, expertise, and teaching practices significantly influence student learning experiences and outcomes (Biggs & Tang, 2011). Shulman (1987) emphasized that faculty with both subject-matter expertise and strong pedagogical skills are better equipped to design and deliver effective instruction, leading to deeper student engagement and understanding. Continuous professional development for academic staff is crucial for integrating innovative teaching methods, adapting to technological advancements, and aligning with evolving academic and industry standards (Knapper & Cropley, 2000). Research also highlights the importance of staff-student interaction in fostering active learning, critical thinking, and emotional support all vital for student success (Chickering & Gamson, 1987). Furthermore, academic staff contribute to program quality through research that informs curriculum development and ensures course content remains relevant to current disciplinary trends (Brew, 2006). Institutions with clear policies on recruitment, evaluation, and professional growth for academic staff are better positioned to maintain high standards of teaching and learning (Devlin & Samarawickrema, 2010). In short, the competence and commitment of academic staff in various dimensions are central to achieve and sustain quality in higher education.

Quality assurance in GETP's facilities and technology

Good facilities and technology are essential for quality tertiary training. Having suitable infrastructure regarding classrooms, labs, libraries, even recreational spaces offers a supportive environment for effective teaching and learning (Temple, 2008). Modern facilities equipped with up-to-date technology enable interactive and innovative teaching methods, like blended learning and virtual simulations, which boost student engagement and understanding (Garrison & Vaughan, 2008). Plus, reliable technology and digital resources support self-directed learning and collaboration, both of which are crucial for developing 21st-century skills (Brown, 2012). It is widely accepted that institutions with well-maintained facilities and current technology tend to attract and retain both students and faculty, which helps build a strong reputation and ensures program quality (Kuh & Hu, 2001). Additionally, integrating technology into teaching, learning and assessment ensures that programs remain relevant to the evolving demands of the labor market (Laurillard, 2012). Periodic evaluations of facilities and technological resources are necessary to address emerging educational needs and sustain quality standards (Oblinger & Oblinger, 2005). Tran (2024) describes modern Vietnamese students as digital natives skilled in technology and inclined toward self-directed learning. However, they still prefer traditional classroom settings over online interaction in English courses. Therefore, investing in state-of-the-art facilities and technology is indispensable for delivering high-quality higher education.

Lecturers' Perceptions

Lecturers play a central role in implementing and maintaining quality assurance systems in higher education institutions. Their perceptions of QA systems can significantly influence the effectiveness and sustainability of these systems (Newton, 2000). Positive perceptions are often

linked to a sense of ownership and involvement in QA processes. Conversely, skepticism or resistance may arise when lecturers view QA as overly bureaucratic or disconnected from teaching realities (Harvey, 2004).

Studies show that lecturers who are actively engaged in QA activities tend to recognize their value in improving teaching practices and student outcomes. For instance, Tam (2001) found that transparent communication and regular feedback mechanisms fostered greater trust and cooperation among academic staff. Similarly, Owlia and Aspinwall (1996) emphasized the importance of aligning QA goals with lecturers' professional development needs to enhance buy-in and participation.

It can be conluded that while existing literature emphasizes the vital role lecturers play in the success of QA systems, most studies have focused on institutional policies or administrative perspectives, with limited attention to how lecturers themselves perceive and engage with QA processes in practice. Although scholars such as Newton (2000) and Harvey (2004) highlight the impact of lecturers' attitudes on the effectiveness of QA, few studies in the Vietnamese context have explored this issue, particularly within GETPs. Additionally, there is a lack of research examining how QA initiatives align with lecturers' professional needs and how their involvement influences program outcomes. These gaps highlight the need for a context-specific investigation into lecturers' perceptions of QA in Vietnamese higher education. Therefore, the present study aims to explore how lecturers perceive the QA system for GETPs, the challenges they encounter, and the practices they value.

Research Questions

To fulfill the purpose of the study, the survey sought to answer the following research question:

What are lecturers' views on the effectiveness of the quality assurance system of General English program?

Methods

Pedagogical Setting & Participants

The two educational institutions surveyed share a common feature: both have been accredited by the Ministry of Education and Training or the Southeast Asian University Network (AUN). Additionally, many of their training programs have met accreditation standards set by both national and international agencies.

Both institutions place significant emphasis on the quality of their GETPs, a concern reflected in employer feedback gathered during interviews with assessors. GE is a critical issue, as many students face challenges graduating due to their inability to obtain an English certificate or communicate effectively in English in professional environments.

Students' limited English proficiency not only restricts their career opportunities but also poses a significant obstacle to the internationalization of higher education. This limitation affects their ability to integrate into and compete within a globalized context.

Table 1

Overview of General English (GE) Training Programs

Institution	Program Name	Duration	Participants	Lecturers	Level Distribution	Credits	Target Standards
HEI 1	English	180 lessons	10,000 students	45 lecturers	4 levels	12 credits	Level 3/6 (Vietnamese 6-level framework) or B1 (CEFR)
HEI 2	English for International Communication	630 lessons	1,960 students	40 lecturers (25 Vietnamese, 15 foreign)	6 levels	20 credits (earned in last 4 levels only)	

Design of the Study

Mixed methods were employed to collect the data, starting with quantitative data collected from the questionnaire, then the qualitative data was collected through interviews with lecturers basing on emerging issues from the survey data.

Data collection & analysis

The questionnaire was desiged basing on eight quality assurance requirements in chapter 2 of the Circular 17/2021/TT-BGDĐT issued by MOET on June 22, 2021 regarding curriculum standards for higher education levels including (1) program objectives, (2) learning outcomes, (3) recruitment standards, (4) study volume, (5) curriculum structure and content, (6) teaching and assessment, (7) teaching and support staff, (8) facilities, teaching technologies and materials. These contents and assessment criteria in National English Accreditation Program (NEAS) were incorporated to generate specific questions in the questionnaire of the study.

After processing the quantitative data to identify interesting or problematic issues, an interview was conducted with 12 lecturers to further discuss these issues. This helped the researcher gain a deeper understanding of the challenges associated with quality assurance and allowed for the suggestion of amendments to improve the implementation of the quality assurance system more effectively. Document analysis was also integrated into the research, focusing on sources such as official notices from institutional websites, course specifications, and training program materials.

Lecturers from HEI1 were coded as L1-1, L1-2, L1-3, L1-4, L1-5, L1-6, L1-7, L1-8, L1-9, L1-10, L1-11, and L1-12. Similarly, lecturers from HEI2 were coded as L2-1, L2-2, L2-3, L2-4, L2-5, L2-6, L2-7, L2-8, L2-9, L2-10, L2-11, and L2-12. In these codes, the first letter “L” stands for “lecturer,” the first digit represents the HEI, and the second digit indicates the lecturer's order within their institution. The data were processed and analyzed for the reliability using Cronbach's alpha, yielding the following results:

Table 2

The questionnaire reliability

Question Groups	Question	Cronbach's Alpha Coefficient	Number of observed variables in the group	Total variable correlation coefficient
1	QA in the GE training program objectives	0.923	5	≥ 0.3
2	QA in the outcomes of the GE Program	0.878	4	≥ 0.3
3	QA in the entrance requirement of the GE program	0.690	3	≥ 0.3
4	QA in the structure and content of the GE program	0.933	4	≥ 0.3
5	QA in the learning volume of the GE program	0.764	4	≥ 0.3
6	QA in teaching methods and assessment of learning outcomes of GE subjects	0.863	6	≥ 0.3
7	Lecturers	0.550	3	≥ 0.3
8	QA in facilities, technology, and learning materials	0.947	4	≥ 0.3
9	External factors affecting the internal quality assurance of the school's General English students	0.693	5	≥ 0.3
10	Internal factors affecting the internal quality assurance of the school's General English training activities	0.680	5	≥ 0.3

Results/Findings

This section summarizes the results of the lecturer survey derived from the quantitative questionnaire (n = 44 in HEI1 and n = 34 in HEI2). Lecturers evaluated the quality assurance (QA) of the General English Teaching Program (GETP) at their respective institutions across multiple key criteria relevant to the QA framework in higher education. These criteria include QA in the objectives of the GE training program, QA in the expected learning outcomes, QA in the entrance requirements, QA in the structure and content of the curriculum, and QA in the learning volume assigned for the program. Additionally, the survey explored QA in teaching methods and the assessment of learning outcomes, which are critical for ensuring effective delivery and student achievement. The roles and perceptions of lecturers, as central stakeholders in QA implementation, were also assessed. Further, QA in the availability and effectiveness of facilities, technology, and learning materials was examined, reflecting the importance of infrastructure in supporting teaching and learning quality. The survey also considered both external factors such as policy, accreditation standards, and stakeholder expectations and internal factors such as institutional leadership, management practices, and staff involvementthat may influence the internal quality assurance of the GE training activities.

To complement and triangulate the survey data, the study also included document analysis of GETP materials and curriculum-related evidence, as well as in-depth interviews with selected lecturers from both institutions. These additional methods provide deeper insights into the QA mechanisms in practice and offer a more comprehensive understanding of how various factors contribute to or hinder the effectiveness of the QA systems in the GETPs at HEI1 and HEI2.

Lecturers' perceived QA in GETP's objectives

Table 3

Lecturers' perceived QA in GETP's objectives

		HEI1		HEI2		Summary of consent levels		
		Mean	**Sd.**	**Mean**	**Sd**	**Mean**	**Sd**	**Order**
OBJ1	The program has clear goals.	4.11	1.10	4.47	0.94	4.29	1.02	5
OBJ2	The program goal is to prepare students for future work skills.	3.97	1.02	4.11	0.92	4.04	0.97	4
OBJ3	The program goal is to improve students' self-learning ability.	4.0	1.09	4.07	0.90	4.03	0.99	4
OBJ4	The program content is designed to achieve the set goals.	4.02	1.02	4.15	0.88	4.08	0.95	4
OBJ5	The courses in the GE program are cohesive.	3.97	1.08	4.19	0.91	4.08	0.99	4
	General Training	**4.014**		**4.19**		**4.10**		**4**
	t-test results	t = -2.001				*Sig.=0.050*		

Table 3 shows that lecturers rated the quality assurance of General English (GE) program objectives more favorably at HEI2 (M = 4.19, SD = 0.91) than at HEI1 (M = 4.01, SD = 1.06), indicating not only higher overall satisfaction but also greater consistency among respondents at HEI2. Among the five criteria, the objective "training program with clear goals" (OBJ1) received the highest rating at both institutions, particularly at HEI2 (M = 4.47), suggesting strong institutional clarity in goal-setting. Other objectives—including the development of work-related skills (OBJ2), promotion of self-directed learning (OBJ3), alignment of content with goals (OBJ4), and cohesion across subjects (OBJ5)—were also rated positively. However, HEI2 consistently outperformed HEI1 across all items.

A t-test confirmed a statistically significant difference between the institutions (t = -2.001, df = 60.198, p = 0.050), with Levene's Test (F = 7.619, p = 0.007) indicating unequal variances. Supporting qualitative data from document analysis revealed that while both HEIs articulate their program objectives in official materials, HEI2's statements are more detailed, measurable, and explicitly aligned with practical competencies and academic outcomes. This suggests a more systematic approach to ensuring quality through well-defined and actionable goals.

In summary, both the quantitative and qualitative findings indicate that HEI2 demonstrates stronger quality assurance practices in setting GE program objectives. These include clearer articulation, better alignment with real-world skills, and stronger internal coherence. HEI1 may enhance its QA efforts by benchmarking against HEI2's more structured and outcome-oriented approach to defining program objectives.

Lecturers' perceived QA in GETP's learning outcomes

Table 4

Lecturers' perceived QA in GETP's learning outcomes

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Aver**	**Sd**	**Level**
LO1	The learning outcomes are realistic and achievable.	3.88	0.94	4.17	0.45	4.02	0.69	4
LO2	Students are evaluated through at least one indicator (project, presentation, diary, workbook).	3.95	1.01	4.20	0.59	4.07	0.80	4
LO3	The learning outcomes are measured to track student progress.	3.70	1.06	4.35	0.48	4.02	0.77	4
LO4	The learning outcomes are always updated to meet the needs of employers.	3.63	1.12	4.44	0.53	4.03	0.82	4
	Average	**3.79**		**4.27**		**4.03**		**4**
	t-test results	*t=-3.342*				*Sig.=.001*		

Table 4 presents that the overall quality assurance level at HEI2 was 4.29, higher than HEI1's score of 3.79, with standard deviations of 0.51 and 1.03, respectively, indicating greater consensus among respondents at HEI2.

For specific criteria, the item "Always updated to meet the needs of employers" (LO4) received the highest rating at HEI2, with a mean of 4.44, while it was rated the lowest at HE1, at 3.63. Similarly, the criterion "Students' progress is measured to track student progress" (L2-3) was highly rated at HEI2 with an average of 4.35, compared to 3.70 at HEI1. Additionally, the criteria "Realistic and achievable performance" (LO1) and "Feedback assessed through at least one indicator" (LO2) had higher mean at HEI2 than HEI1, reflecting HEI2's effectiveness and practicality in curriculum design. Overall, HEI2 demonstrated higher results than HEI1 across all criteria, featuring its superiority in quality assurance. The t-test results further confirm a statistically significant difference between the mean scores of the two institutions, with HEI2 having a higher mean score. To further explain for the difference, the document study was conducted, which show that the curriculum information for HEI1 and HEI2 emphasizes the role of lecturers in shaping the position and importance of the subjects within the program. Course objectives are articulated using levels of cognitive capacity, skills, and attitudes. Both curricula clearly outline expectations for the four core skills: Listening, Speaking, Reading, and Writing. At HEI1, L1-7 explained that the course outcomes include "understanding the main points of clear, standardized information on familiar professional topics such as employment, technology, tourism, and history," and "listening to short, simple monologues or conversations using high-frequency vocabulary in professional contexts." L1-9 added that students are also expected to "write CVs, formal emails requesting information, and reviews of films, books, websites, and products." In addition, learners should be able to "engage in simple, everyday communication

tasks requiring direct exchanges of information" and "demonstrate continuous progress in English learning throughout the course."

At HEI2, L2-4 noted that students should "understand the main points of standard speech on common problems encountered in work, school, or entertainment" and "grasp the gist of radio or television programs on relevant topics." According to L2-7, the curriculum also expects learners to "communicate effectively in most travel-related situations, engage in unprepared conversations on familiar topics, and describe experiences, dreams, or stories in a simple but connected manner." As highlighted by L2-6, reading and writing outcomes focus on the ability to "understand texts related to work or daily life, comprehend descriptions of events, emotions, and desires in personal correspondence," and "write paragraphs describing processes, expressing opinions, narrating events, or interpreting data."

Overall, the quantitative data and all interviewed lecturers verified that the training programs were reasonable and contributed to improving the overall quality of education.

Lecturers' perceived QA in GETP's entry requirements

Table 5

Lecturers' perceived QA in GETP's entry requirements

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Level**
Entry1	An effective placement test to place students in appropriate classes.	4.04	1.07	4.17	0.45	4.10	0.76	4
Entry2	Students are enrolled in courses with a level appropriate to their language proficiency.	3.18	0.72	3.55	0.84	3.16	0.78	3
Entry3	Placement tests are periodically adjusted to be up-to-date.	3.93	1.14	4.01	0.97	3.97	1.05	4
	General Training	**3.71**		**3.91**		**3.74**		**4**
	t-test results	*t.=-2.676*				*Sig. =.009*		

Table 5 presents the lecturers' evaluations of the level of quality assurance in the entry requirements of the GETPs at two educational institutions. The results indicate differences between the two institutions. For the general criterion "Ensuring the quality of input for the GE subject program," HEI2 received a higher mean score and a lower standard deviation (SD) (0.75 compared to 0.97), reflecting greater uniformity in assessment.

In the specific criteria, HEI2 consistently outperformed HEI1. For Entry1, HEI2 had a higher mean score (4.17 compared to 4.04) and a lower Sd (0.45 compared to 1.07), indicating a more stable and favorable evaluation of the effectiveness of the placement test. For Entry2, the mean scores were lower for both institutions, with HEI1 scoring 3.18 and HEI2 scoring 3.55. The Sds were 0.72 and 0.84, respectively. For Entry3, which evaluates the periodic update of the placement test, both institutions achieved high mean scores, with HEI2 again leading (4.01 vs. 3.93). However, the Sd at HEI1 (1.14) was significantly higher than at HEI2 (0.97).

The t-test results indicate that HEI2 had a significantly higher average score than HEI1. Research on the structure of the GE entrance placement tests at the two institutions reveals that the tests assess three skills: Listening, Reading, and Writing, with no Speaking component.

According to management staff, the inclusion of three skills is sufficient for accurate student placement, as adding a Speaking component would unnecessarily complicate the exam process.

When evaluating Entry 2, lecturer L1-3 noted,

> "Students are not always enrolled in courses suitable for their language level. Many students struggle in English 1 because they only studied English for three years in high school or not at all. Students scoring below 4 are placed in English 1, but perhaps the school should introduce more basic English classes to support students with scores ranging from 0 to 3."

Lecturers' perceived QA in GETP's structure and content

Table 6

Lecturers' perceived QA in GETP's structure and content

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Level**
ST1	The courses are designed to meet the learning needs of students.	3.63	1.08	4.02	0.62	3.82	0.85	4
ST2	Courses are designed based on developments in language and technology teaching methods.	3.65	0.88	4.02	0.62	3.83	0.75	4
ST3	Each course has specific, and measurable goals.	3.88	1.08	4.29	0.46	4.08	0.77	4
ST4	Curriculumn materials effectively support lecturers in planning and implementing lessons.	3.93	1.06	4.14	0.65	4.03	0.85	4
	Average	**3.77**		**4.12**		**3.94**		**4**
	t-test results	t = -2.077			*Sig.=.042*			

However, lecturers at HEI1 note some challenges regarding student preparedness and the results in table 6 show that HEI2 received higher ratings than HEI1 across all criteria. HEI2 achieved a higher mean score and lower standard deviation, suggesting a greater level of agreement among respondents.

The overall average score for HEI2 was 4.12, notably higher than HEI1's 3.77. Among the criteria, ST3 ("Each course has specific, measurable goals") received the highest mean score at both institutions, particularly at HEI2 (4.29). Similarly, the criterion for curriculum materials showed strong performance, with high mean scores and good consistency across both campuses. Overall, HEI2 is regarded as superior in program structure and content, with clear advantages in specific criteria. The results of Levene's Test ($F = 7.029$, $p = 0.010$) confirms that HEI2's average score is significantly higher than HEI1's.

Both HEI1 and HEI2 strive to enhance their GETP to better meet students' needs. They incorporate technology each semester, including LMS exercises, e-workbooks, and video games, to support English learning. The objectives of each course are clearly defined and measurable, with carefully selected teaching materials and structured lesson plans ensuring consistency in content delivery and progress throughout the program.

> "The current courses are designed for students with an A2-level foundation according to the CEFR. Students without this background face significant difficulties, resulting in a

high failure rate in English 1, especially among engineering students." (L1-5)

"The courses reflect advancements in language teaching methods and technology, but the ability to apply technology varies among lecturers. Younger lecturers in their 30s and 40s tend to use technological tools more effectively than their older counterparts." (L1-7)

Overall, HEI2 was rated more favorably and showed greater consistency in lecturer opinions regarding entry quality assurance criteria.

Lecturers' perceived QA in GETP's learning volume

Table 7

Lecturers' perceived QA in GETP's learning volume

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Average**	**Sd**	**Level**
VOL1	There are enough courses at each level to meet the real needs of students.	3.77	1.29	4.02	0.62	3.89	0.95	4
VOL2	Each level must have 200 hours of class contact and supervised learning.	3.50	1.26	3.94	1.13	3.72	1.19	4
VOL3	There is an e-learning system to support General English learning.	3.47	1.04	4.47	0.78	3.97	0.91	4
VOL4	Students are given the opportunity to expand their language learning outside of the classroom.	3.29	1.35	4.11	0.53	3.70	0.94	4
	General Training	**3.51**		**4.14**		**3.82**		**4**
	t-test results	t = -4.906				Sig. =.000		

Table 7 shows that HEI2 outperformed HEI1 across all criteria, with higher mean scores and lower SDs, indicating stronger and more consistent evaluations. HEI2 scored 4.14 for "Ensuring learning volume" compared to HEI1's 3.51. The highest-rated item at HEI2 was VOL3 (e-learning system) at 4.47, while VOL4 (extracurricular language learning) showed the largest SD gap—1.35 at HEI1 vs. 0.53 at HEI2.

Lecturers at HEI1 noted limited course offerings: "Many students have to wait until their second or third year to take English classes" (L1-1, L1-3), delaying graduation. HEI1 offers 45 face-to-face and 90 self-study hours per level, while HEI2 provides 105 hours in class. Both fall short of the CEFR's 200-hour standard due to credit restrictions: "We can only allocate a maximum of 105 hours per level" (L2-2).

HEI1 has added two levels to reduce delays, using e-workbooks, speaking tasks, and group projects to enhance self-study (L1-4). At HEI2, students practice English in real-life settings: "They interview foreigners on topics like culture and tourism, after drafting questionnaires and gaining approval" (L2-4).

Lecturers' perceived QA in GETP's teaching and assessment methods

Table 8

Lecturers' perceived QA in GETP's teaching and assessment methods

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Average**	**Sd**	**Level**
T1	Lecturers utilizes student-centeredness to maximize engagement.	3.84	1.05	4.29	0.62	4.06	0.83	4
T2	Assessment methods are diverse, including initial, procedural and summary assessments.	4.02	1.06	4.38	0.93	4.20	0.99	4
T3	Lecturers use feedback and editing techniques to maximize student learning and engagement.	3.77	1.00	4.32	0.58	4.04	0.79	4
T4	Lecturers integrate technology to support effective learning.	4.15	0.71	4.23	0.55	4.19	0.63	4
T5	Lecturers arrange lessons and activities in alignment with the the CLOs.	3.97	0.40	4.29	0.46	4.13	0.43	4
T6	Lecturers have teaching strategies suitable for the objectives and levels	4.15	0.88	4.38	0.55	4.26	0.71	5
	General Training	**3.98**		**4.31**		**4.14**		**4**
	t-test results	t = -2.526				Sig.=.014		

The table presents lecturers' perceptions of quality assurance (QA) in the teaching and assessment methods of General English Teaching Programs (GETPs) at two higher education institutions, HEI1 and HEI2. Overall, lecturers from both institutions rated the QA practices positively, with HEI2 consistently receiving higher mean scores across all items. The general average score was 4.14, reflecting a high level of perceived QA, with HEI2 scoring 4.31 compared to HEI1's 3.98. The t-test result ($t = -2.526$, $p = 0.014$) indicates a statistically significant difference between the two institutions.

Among the six criteria, the highest-rated item was T6—"Lecturers have teaching strategies suitable for the objectives and levels"—with a combined mean of 4.26, rated particularly high at HEI2 (4.38). This reflects strong alignment between instructional methods and student needs. T2 and T5, which assess the diversity of assessment methods and the alignment of lessons with course learning outcomes (CLOs), also received high ratings from both institutions. The lowest-rated item at HEI1 was T3—"Use of feedback and editing techniques"—with a mean of 3.77, compared to 4.32 at HEI2, suggesting more effective feedback practices at HEI2.

Standard deviations were generally lower at HEI2, indicating more agreement among its lecturers, especially in areas such as the use of technology (T4) and student-centeredness (T1). These results suggest that HEI2 lecturers not only perceive stronger QA in teaching and assessment but also demonstrate greater consistency in their evaluations.

In summary, the data highlights HEI2's superior performance in implementing QA measures in teaching and assessment. This underscores the need for HEI1 to review and possibly adopt effective practices from HEI2, particularly in feedback techniques, alignment with CLOs, and the integration of student-centered strategies.

Lecturers' perceived QA in GETP's academic staff

Table 9

Lecturers' perceived QA in GETP's academic staff

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Average**	**Sd**	**Level**
VII	**QA in academic staff**							
LEC1	Lecturers have qualifications that are suitable for the requirements of undergraduate GE teaching.	4.38	0.65	4.55	0.50	4.46	0.57	5
LEC2	Lecturers are assigned to teach at different levels based on their experience and training.	4.00	0.71	4.29	0.67	4.14	0.69	4
LEC3	Lecturers can participate in training and continuous professional development courses to improve their GE teaching capacity.	3.79	1.26	4.38	0.55	4.08	0.90	4
	General Training	**4.05**		**4.40**		**4.22**		**5**
	t-test results	t = -2.720				Sig.=.008		

Table 9 summarizes that HEI2 outperformed HEI1 across all criteria, with higher mean scores and lower standard deviations, indicating greater stability in evaluations. Regarding the general assessment of quality assurance, HEI2 scored 4.40 (sd = 0.57) compared to HEI1's 4.05 (sd = 0.87). Levene's Test (F = 0.935, p = 0.337) supports the equal variance assumption. The t-test results (t = -2.720, df = 76, p = 0.008) indicate a statistically significant difference between the two groups, with HEI2 showing a higher mean difference (-0.35116, 95% CI: [-0.60829, -0.09402]). Overall, HEI2 demonstrated superior and more consistent quality assurance in academic staff compared to HEI1.

When commenting on the criterion "Lecturers can participate in training and retraining courses to improve their GE teaching capacity," lecturers at HEI1 said:

> "At the beginning of each semester, the head of the department has integrated in the orientation meeting a short teaching workshop such as training on the use of electronic workbooks. However, I personally feel that the department and faculty should have a master plan on organizing training for lecturers to teach GE. For example, faculties and departments need to collect lecturers' opinions on training needs, then periodically organize internal seminars close to the needs of lecturers with a frequency of 4 times a year to help ensure the same teaching quality in all classes." (L1-8)

This feedback highlights the need for a structured and proactive approach to professional development. A master plan that incorporates lecturer feedback and organizes training sessions regularly would address specific needs and promote consistent teaching quality.

> "Professional development activities need to be more focused. I think it is necessary to increase peer observation so that lecturers can learn from each other, so that the teaching method will be more uniform." (L1-6)

Encouraging peer observation is an excellent suggestion, as it fosters peer learning and helps harmonize teaching methods across classes. This can enhance both teaching quality and student

experience.

> "Many of the lecturers are dynamic in adopting modern teaching methods, which makes the classroom lively and students have the opportunity to practice the language in a fun and highly applicable context, while many classes are taught in a rather traditional, dull classroom atmosphere. I think the head of the department should increase the number of comments so that lecturers can change their teaching methods to be more exciting and effective. Attendance should be a priority to develop the teaching team at the department." (L1-2)

The observation underlines the disparity in teaching approaches. Increasing feedback and encouraging dynamic, modern methods across the department would benefit both lecturers and students. Prioritizing attendance at training and development sessions is key to fostering a cohesive teaching team.

At HEI2, the lecturers said:

> "Lecturers are informed about English teaching seminars held in Ho Chi Minh City to arrange attendance, such as the annual VUSTESOL, which is held free of charge for the community of English lecturers." (L2-7)

Sharing information about local seminars like VUSTESOL is an effective way to provide accessible professional development opportunities.

> "Webinars on English teaching are the place where the university's lecturers participate the most for many reasons, such as being free to attend, lecturers not having to travel, only needing to register and connect with a laptop, and offering a variety of topics to serve the diverse professional development needs of lecturers." (L2-8)

Webinars offer flexibility and accessibility, catering to diverse professional development needs. Institutions could further encourage participation by sharing a calendar of relevant webinars.

> "Lecturers join some communities such as VietTESOL, to attend and review diverse sources of webinars." (L2-9)

Participation in professional communities like VietTESOL provides lecturers with ongoing access to resources and peer support, which are vital for continuous development.

> "Cambridge and Oxford also regularly offer free online teaching workshops for English lecturers." (L2-10)

Collaborating with global institutions like Cambridge and Oxford offers learning opportunities, enhances the quality of professional development, and ensures exposure to international practices.

> "The General English Training Program has been accredited by NEAS, and every month lecturers have the opportunity to attend CPD (Continuous Professional Development) courses to accumulate professional development points as well as update modern teaching methods." (L2-6)

Regular CPD courses accredited by NEAS provide lecturers with structured, high-quality training, fostering continuous improvement in teaching methods.

To recap, at HEI1, the focus should be on implementing a structured professional development strategy, increasing interaction among lecturers, and encouraging the adoption of innovative teaching methods. In the meanwhile, at HEI2, the institution effectively utilizes a variety of professional development platforms, such as webinars, workshops, and accredited courses, to

support lecturers' growth. Sharing best practices from HEI2 could inspire HEI1 to expand and refine its training initiatives.

Lecturers' perceived QA in GETP's facilities and technology

Table 10

Lecturers' perceived QA in GETP's facilities and technology

		HEI1		HEI2		Average		
		Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Level**
VIII	**QA in facilities, technology, and learning materials**							
FAC1	Facilities are invested to support the quality of GET.	3.36	1.16	4.47	0.50	3.91	0.83	4
FAC2	Technology is invested to support the quality of GET.	3.22	1.13	4.38	0.55	3.80	0.84	4
FAC3	Learning materials are invested enough to support the quality of GET.	3.68	1.28	4.41	0.49	4.04	0.88	4
FAC4	Designing teaching spaces creates conditions for students to actively participate in lesson development.	3.18	1.36	4.29	0.46	3.73	0.91	4
	General Training	**3.36**		**4.38**		**3.87**		**4**
	t-test results	t = -5.528				Sig.=.000		

Table 10 shows that HEI2 had a significantly higher mean than HEI1 on all factors, with an overall rating of 4.38 compared to 3.36. At the same time, HEI2's sd is lower, indicating a higher consensus among lecturers. Factors such as "facilities" and "learning materials" at HEI2 all received positive evaluations (4.47 and 4.41 respectively), reflecting a better and more effective investment than HEI1. The results of Levene's Test ($F = 36.745$, $p = 0.000$) suggests that HEI2 had significantly higher mean than HEI1.

At HEI1 the classroom used for teaching GE is equipped with two ceiling loud speakers, two air conditioners, long tables for 2 students to sit and difficult to turn, a chalk board, a projector and a canvas screen, two ceiling speakers, microphone plugs, internet cable for lecturers, weak wifi system. Lecturers bring their own laptops, speaker cables, and personal microphones to plug in. Commenting on facilities for GE teaching, lecturers at HEI1 stated:

> "Tables and chairs are not suitable for organizing English teaching activities. The benches and tables are close to each other, making it very difficult for lecturers to organize group activities. Most lecturers can only let students work in pairs" (L1-9)
>
> "The walls between the layers of sound insulation are not very good. Many times when the lecturer in the next room uses a speaker or microphone, my class itself is greatly affected. The audio and the voice of the lecturer in the next room drowned out the sound of my class." (L1-2)
>
> "The classroom space is not suitable for organizing English teaching. Sometimes I want to let the students stand up for questions and answers for practical role-playing, I find it difficult because I can't get a narrow classroom, I can't let students move around." (L1-10)
>
> "There are a lot of resources available online to teach English. However, the wifi is not covered enough for lecturers to apply games or activities that need to use the network." (L1-6)

At HEI2 shows that classrooms are equipped with removable chairs and individual removable tables, making it easy to move around in pairs and groups. The classroom is equipped with very good soundproof walls, modern air conditioning, a projector, a canvas screen, a desktop computer connected to the internet by cable, a computer pre-installed with the iTools of the textbook, and a wifi system that is strong enough for students to carry out online English learning activities. Commenting on the facilities at HEI2, the lecturers said:

> "I am satisfied with the facilities at the school. When I teach, I just need to compile more documents and send emails to me personally. When I get to class, I log in to the available desktop computer to download the lessons. I go to teach very lightly, I don't need to carry a lot of equipment like when I teach at other schools." (L2-3)
>
> "I love the classroom space at school. There's a large enough space in front of the classroom for me to let students do some activities that require mobility." (L2-5)
>
> "The teaching space here includes space outside the classroom. There are foreign lecturers here, so students have the opportunity to communicate when meeting them in the corridors or common areas of the school" (L2-1)

To recap, HEI2 demonstrates a clear advantage in facilities and resources, significantly enhancing teaching and learning experiences, while HEI1 struggles with inadequate infrastructure that limits teaching methods and activities.

Discussion

Quality assurance in GETP's objectives

The findings on the GETPs' objectives highlight a consistent emphasis on clarity and alignment across institutions. Document analysis reveals that both HEI1 and HEI2 articulate specific and practical goals aimed at equipping students with the necessary language skills for academic and professional contexts. HEI1 focuses on enabling students to use language in simple, professional exchanges, while HEI2 extends these goals to include understanding speech, handling real-world situations, and developing academic literacy. Despite these variations, both institutions prioritize practical application and alignment with students' broader learning needs. This consistency resonates with the educational theory suggesting that clear and well-defined objectives provide a strong framework for curriculum design and implementation (Anderson & Krathwohl, 2001).

Interviews with lecturers reinforce the importance of these objectives, as they provide clear descriptions for program activities. This supports Biggs and Tang's (2011) assertion that objectives foster consistency and coherence in program delivery. Moreover, the emphasis on designing objectives based on existing programs reflects Tyler's (1949) principle of using objectives as benchmarks for improvement and success evaluation.

A notable distinction, however, lies in the broader scope of HEI2's objectives, which include academic literacy and prepare students with specialized subjects. This aligns with research advocating for SMART objectives to address specific institutional and student needs (Doran, 1981; O'Neill, 2020). Overall, the findings underscore that while variations exist, the shared commitment to clear and practical objectives contributes to achieving high standards of quality in the GETPs.

Quality assurance in GETP's learning outcomes

Contrasting GETP learning outcomes at HEI1 and HEI2 experiences notable differences. HEI2 consistently outperformed HEI1, demonstrating superior curriculum design and implementation, particularly in updating outcomes to meet employer needs and tracking student progress (Biggs & Tang, 2011). While both institutions align course objectives with cognitive skills, practical abilities, and attitudinal development (Anderson & Krathwohl, 2001), HEI1 showed lower scores and greater variability. Statistically significant differences identified HEI2's advantage in quality assurance, particularly its alignment with measurable and transparent learning outcomes (Harden, 2002). HEI2's more effective design and implementation serve as a model for aligning learning outcomes with institutional and the requirements of labor markets.

Quality assurance in GETP's entry requirements

The analysis of entry requirements for GETPs at HEI1 and HEI2 reveals significant differences in quality assurance. HEI2 consistently outperforms HEI1, demonstrating a more cohesive and robust approach to student placement (Yorke & Longden, 2004). HEI2's higher scores and greater consistency across specific criteria, like placement test effectiveness and periodic updates, align with research emphasizing the importance of clear, aligned entry requirements for student success and retention (Kuh et al., 2006; Galloway, 2009). However, qualitative feedback reveals challenges. Particularly, HEI1 where there is misalignment between student readiness and placement, especially for students with minimal English background, points out the need to balance accessibility and quality (Smith & Naylor, 2001). While both institutions make efforts to ensure quality, HEI2's more comprehensive and consistent approach appears more effective in fostering academic fit.

Quality assurance in GETP's structure and content

A comparison of GETP structure and content at HEI1 and HEI2 reveals key differences. HEI2's higher scores and greater consistency reflect superior performance (Barnett, 2000; Biggs & Tang, 2011). HEI2 excelled in clear course goals and relevant curriculum materials. HEI1 faces challenges with student preparedness and varying lecturer technological proficiency, echoing the need for continuous program review (Fry et al., 2008). For example, HEI1's struggled with students lacking A2-level English skills impact success rates, focusing the importance of perceived program purpose and attainability (Merrill, 2002). While both institutions aim for improvement, HEI2's structure and content more closely align with effective program design principles, better preparing students for current demands.

Quality assurance in GETP's learning volume

The findings on learning volume assurance for GETPs at HEI1 and HEI2 reveal notable contrasts in their approaches and outcomes. HEI2 offers diverse learning opportunities, including real-world application projects, aligning with research on balanced learning activities (Biggs & Tang, 2011). HEI1 faces challenges with insufficient courses, impacting enrollment and graduation, and struggles to meet CEFR-recommended learning hours despite adjustments. These differences reflect varying approaches to workload comparability and student well-being (Adam, 2004; Kember, 2004). While HEI1 relies heavily on self-study, HEI2's structured approach appears more effective in ensuring consistent student progress and program quality.

Quality assurance in GETP's teaching and assessment methods

The findings on teaching methods and assessment practices in the GETPs at HEI1 and HEI2 highlight clear differences and align with previous studies on effective educational strategies.

HEI2 consistently outperformed HEI1, achieving a higher mean score and demonstrating greater consensus, as indicated by lower standard deviations. Notably, the criteria for diversity in assessment methods and appropriate teaching strategies received the highest ratings at both institutions (4.38), reflecting a shared emphasis on key elements of quality teaching. These results align with Prince's (2004) assertion that diverse and active pedagogical approaches, such as problem-based and experiential learning, enhance student engagement and critical thinking.

A significant disparity observed in the criterion for effective feedback and editing brings out HEI2's strength in providing formative feedback, consistent with Sadler's (1989) emphasis on its role in fostering student improvement. Additionally, the statistical results confirm a significant overall difference between the two institutions, reinforcing HEI2's superiority in teaching and assessment practices.

These findings align with Biggs' (2003) concept of constructive alignment, which integrates teaching methods and assessment practices with learning outcomes to ensure coherence and quality. However, the variability in performance suggests areas for HEI1 to improve, particularly in implementing diverse and inclusive assessment methods (Gibbs & Simpson, 2004) and providing timely feedback. Overall, HEI2 exemplifies the thoughtful integration of innovative teaching and assessment practices, which is crucial for maintaining high-quality tertiary education, as emphasized in the literature.

Quality assurance in GETP's academic staff

The findings on academic staff quality assurance in the GETPs at HEI1 and HEI2 highlight differences in institutional practices and align with prior studies on faculty development. HEI2 outperformed HEI1 across all criteria, with a higher mean score and lower standard deviations, indicating greater consistency in evaluations. The t-test results confirm a statistically significant difference, emphasizing HEI2's superior approach to academic staff development. This aligns with Devlin and Samarawickrema's (2010) assertion that clear policies on faculty growth are crucial for maintaining high standards.

At HEI2, lecturers benefit from a variety of professional development opportunities such as accredited CPD courses, webinars, and participation in professional communities like VietTESOL. These initiatives provide accessible and diverse training options, which enables lecturers to stay updated with modern teaching methods and global standards, as emphasized by Knapper and Cropley (2000). For example, regular webinars and workshops by Cambridge and Oxford ensure exposure to innovative practices, while NEAS accreditation emphasizes the institution's commitment to quality assurance.

In contrast, HEI1 faces challenges such as a lack of structured professional development plans and inconsistencies in teaching practices. Feedback from HEI1 lecturers suggests a need for systematic training, such as peer observation and frequent workshops tailored to lecturers' needs. These suggestions align with Shulman's (1987) emphasis on equipping faculty with both subject-matter knowledge and pedagogical skills to enhance teaching effectiveness.

To recap, while HEI2 exemplifies a robust and proactive approach to academic staff development, HEI1 could benefit from adopting similar strategies, such as structured training plans and collaborative learning practices, to improve teaching quality and consistency. Sharing best practices from HEI2 is believed to encourage HEI1 to better support its lecturers, ultimately enhancing the quality of its GE program.

Quality assurance in GETP's facilities and technology

The findings on facilities and technology in the GETPs at HEI1 and HEI2 show great contrasts. With a higher overall mean score and a statistically significant mean difference, HEI2 demonstrates a clear advantage. HEI2's consistent lecturer evaluations point to the high quality of its facilities (Temple, 2008; Garrison & Vaughan, 2008). In the meanwhile, HEI1 faces challenges with inadequate classroom layouts, poor soundproofing, and weak wifi, hindering interactive teaching (Oblinger & Oblinger, 2005). Conversely, HEI2 provides state-of-the-art resources, leading to higher lecturer satisfaction and a conducive environment for dynamic teaching (Kuh & Hu, 2001). While HEI1 struggles with limitations, HEI2 leverages its superior facilities. It is implicated that adopting HEI2's practices could significantly improve HEI1's teaching infrastructure.

Conclusion

This study stresses the critical role of quality assurance in the success and sustainability of GETPs in higher education. A comparative analysis of two institutions reveals HEI2 as a model of best practices across multiple dimensions, containing objectives, learning outcomes, entry requirements, learning volume, program structure and content, teaching methods, academic staff, and facilities. Key findings suggest that institutions with clearly defined objectives, measurable learning outcomes, robust entry criteria, and well-structured programs consistently outperform those with less cohesive QA frameworks. HEI2's alignment with international standards like the CEFR, its strategic use of technology, and its investment in academic staff development demonstrate the combined impact of these factors on program effectiveness, student satisfaction, and graduate employability.

However, challenges such as resource limitations, stakeholder misalignment, and resistance to change persist, particularly at HEI1. These issues are consistent with previous research indicating that inadequate funding, lack of human resources, and conflicting interests among stakeholders often hinder the implementation of quality assurance (QA) initiatives in higher education institutions (Harvey & Newton, 2007; Materu, 2007). Moreover, resistance to change can stem from deeply rooted institutional cultures and the perceived threat of QA processes to academic autonomy (Stensaker, 2008). These findings reinforce the need for institutions to adopt best practices from successful models, such as HEI2, by fostering stakeholder collaboration (Santiago et al., 2008), leveraging innovative teaching technologies (Guri-Rosenblit, 2005), and embedding continuous improvement in their QA frameworks (Van Vught & Westerheijden, 1994). Such approaches not only enhance institutional effectiveness but also promote a culture of quality that supports long-term educational development.

This study just focused on the perception of the most important internal stakeholder-lecturers in enhancing and maintaining quality assurance of an academic program. Our following research will focus on the longitudinal impact of QA practices on student outcomes and investigate how scalable strategies can address diverse institutional contexts and constraints.

Acknowledgments

We would like to show our gratitude to lecturers, program directors, the dean, and vice deans at the participating institututes for their intellectual, and spriritual contributions to this research article.

References

Adam, S. (2004). Using learning outcomes: A consideration of the nature, role, application, and implications for European education of employing 'learning outcomes' at the local, national, and international levels. *United Kingdom Bologna Seminar.* https://doi.org/10.1016/j.tate.2004.09.012

Alvesson, M., & Sandberg, J. (2013). Has management studies lost its way? Ideas for more imaginative and innovative research. *Journal of Management Studies*, *50*(1), 128–152. https://doi.org/10.1111/j.1467-6486.2012.01070.x

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

Balabag, V. C., & Candilas, K. S. (2024). Improving College Students' Metacognition through Self-Paced Learning and Direct Instruction in a Remote Learning Modality. *BRU ELT JOURNAL*, *2*(2), 142-157. https://doi.org/10.14456/bej.2024.11

Barnett, R. (2000). Realizing the university in an age of supercomplexity. *Society for Research into Higher Education & Open University Press*.

Barrie, S. C. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education*, *51*(2), 215–241.

Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education*, *41*(3), 221–238. https://doi.org/10.1023/A:1004181331049

Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and Learning in Higher Education: New Trends and Innovations*, *13*(2), 13–17.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). McGraw-Hill Education.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge. https://doi.org/10.4324/9780203964309

Brew, A. (2006). *Research and teaching: Beyond the divide*. Palgrave Macmillan. https://doi.org/10.1057/9780230591904

Brookhart, S. M. (2023). Assessment Literacy in a Better Assessment Future. *Chinese Journal of Applied Linguistics*, *46*(2), 162-179. https://doi.org/10.1515/CJAL-2023-0202

Brown, H. D. (2014). Principles of language learning and teaching (6th ed.). Pearson Education.

Brown, G., & Knight, P. (1994). *Assessing learners in higher education*. Kogan Page.

Brown, M. (2012). Learning spaces: Using the potential of physical and virtual spaces in teaching and learning. In M. Keppell et al. (Eds.), *Physical and virtual learning spaces in higher education* (pp. 41–57). IGI Global. https://doi.org/10.4018/978-1-4666-1149-0.ch003

Carless, D. (2023). Teacher feedback literacy, feedback regimes and iterative change: towards enhanced value in feedback processes. *Higher Education Research & Development*, *42*(8), 1890–1904. https://doi.org/10.1080/07294360.2023.2203472

Chen, Z., & Goh, C. C. M. (2011). Teaching oral English in higher education: Challenges to EFL teachers. *Teaching in Higher Education*, *16*(3), 333–345.

https://doi.org/10.1080/13562517.2010.546527

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *39*(7), 3–7.

Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education*, *11*(1), 25–36. https://doi.org/10.1080/13538320500074915

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

Dang, T. M. T., & Pham, T. Q. (2024). Alignment between Course Learning Outcomes and Assessments: An Analysis within Linguistic Programs at a University in Vietnam. *International Journal of TESOL and Education*, *4*(2), 31-45. DOI: https://doi.org/10.54855/ijte.24422

Dang, T. M. T., & Tong, T. M. H. (2024). Instruction-Giving in EFL Classrooms: A Case of a University in Viet Nam. *International Journal of TESOL & Education*, *4*(2), 46-61. DOI: https://doi.org/10.54855/ijte.24423

Devlin, M., & Samarawickrema, G. (2010). The criteria of effective teaching in a changing higher education context. *Higher Education Research & Development*, *29*(2), 111–124. https://doi.org/10.1080/07294360903244398

Doran, G. T. (1981). There's a SMART way to write management's goals and objectives. Management Review, *70*(11), 35–36.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.

Driscoll, A., & Wood, S. (2007). Developing Outcomes-Based Assessment for Learner-Centered Education: A Faculty introduction. http://www.bchs.edu/files/u41/nt_for_Learner-centered_Education-Weingarten.pdf

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

European Commission. (2015). *ECTS users' guide 2015*. Publications Office of the European Union. https://doi.org/10.2766/87192

Fry, H., Ketteridge, S., & Marshall, S. (2008). *A handbook for teaching and learning in higher education: Enhancing academic practice* (3rd ed.). Routledge.

Galloway, F. J. (2009). Entry requirements and student retention: Balancing quality and access in tertiary education. *Journal of Higher Education Policy and Management*, *31*(3), 245–259.

Gao, X. (2013). The challenges of learning English in higher education: A case study. *Language Teaching Research*, *17*(4), 473–494. https://doi.org/10.1177/1362168813494125

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass. https://doi.org/10.1002/9781118269558

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31.

Government of Vietnam (2016, October 18). Vietnamese qualifications framework, approved

by the prime minister of Vietnam. Decision 1982/QD-TTg. https://chinhphu.vn/default.aspx?pageid=27160&docid=186972

Gu, P. Y., & Lam, R. (2023). Developing Assessment Literacy for Classroom-Based Formative Assessment. *Chinese Journal of Applied Linguistics*, *46*(2), 155-161. https://doi.org/10.1515/CJAL-2023-0201

Guri-Rosenblit, S. (2005). 'Distance education' and 'e-learning': Not the same thing. *Higher Education*, *49*(4), 467–493. https://doi.org/10.1007/s10734-004-0040-0

Hoang, V.V. (2010). The current situation and issues of the teaching of English in Vietnam. 立命館言語文化研究, *22*(1), 7-18.

Harden, R. M. (2002). Learning outcomes and instructional objectives: Is there a difference? *Medical Teacher*, *24*(2), 151–155.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9-34. https://doi.org/10.1080/0260293930180102

Harvey, L. (2004). The power of accreditation: Views of academics. *Journal of Higher Education Policy and Management*, *26*(2), 207-223. https://doi.org/10.1080/1360080042000218267

Harvey, L., & Newton, J. (2007). Transforming quality evaluation: Moving on. In D. F. Westerheijden, B. Stensaker, & M. J. Rosa (Eds.), *Quality Assurance in Higher Education: Trends in Regulation, Translation and Transformation* (pp. 225–245). Springer.

Ho, N. P., & Ha, Y. N. (2025). Evaluation of the Effectiveness of Learning Outcomes from Students' Perspectives towards Research Writing Course: A Case Study. *International Journal of Language Instruction*, *4*(1), 33–49. https://doi.org/10.54855/ijli.25413

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.

Johnson, R., Brown, A., & Smith, J. (2020). Impact of self-paced learning on metacognition in college students: A comparative study. *Educational Psychology Review, 38*(3), 567-582. https://doi.org/10.1007/s10648-020- 09558-2

Kember, D. (2004). Interpreting student workload and the factors which shape students' perceptions of their workload. *Studies in Higher Education*, *29*(2), 165–184. https://doi.org/10.1080/0307507042000190778

Kennedy, D., Hyland, Á., & Ryan, N. (2009). Learning outcomes and competences. In E. Froment et al. (Eds.), *EUA Bologna Handbook: Making Bologna Work*. Raabe Academic Publishers.

Knapper, C. K., & Cropley, A. J. (2000). *Lifelong learning in higher education* (3rd ed.). Routledge.

Kohoutek, J. (2009). Quality assurance in higher education: A case study approach. *Quality in Higher Education*, *15*(2), 133-145. https://doi.org/10.1080/13538320902995764

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature*. National Postsecondary Education Cooperative.

Kuh, G. D., & Hu, S. (2001). The effects of student-faculty interaction on students'

educational gains in college. *Research in Higher Education*, *42*(3), 309–332. https://doi.org/10.1023/1010881521483

Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge. https://doi.org/10.4324/9780203125083

Martin, M., & Stella, A. (2007). *External quality assurance in higher education: Making choices*. UNESCO.

Materu, P. (2007). Higher Education Quality Assurance in Sub-Saharan Africa: Status, Challenges, Opportunities, and Promising Practices. *World Bank Working Paper No. 124*. Washington, DC: The World Bank.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43–59. https://doi.org/10.1007/BF02505024

Ministry of Education and Training (MOET). (2021). *Circular 17/2021/TT-BGDĐT on Regulations for Assessment and Evaluation in General Education Institutions*. Issued June 22, 2021.

Newton, J. (2000). Feeding the beast or improving quality? *Quality in Higher Education*, *6*(2), 153-163. https://doi.org/10.1080/713692740

Nguyen, V. L. (2024). The Impact of Online Resources via Personal Devices on English-Majored Students' Learning Practices. *International Journal of Language Instruction*, *3*(4), 110–118. https://doi.org/10.54855/ijli.24346

Oblinger, D. G., & Oblinger, J. L. (2005). *Educating the net generation*. EDUCAUSE.

O'Neill, G. (2020). Designing effective learning outcomes. *Teaching in Higher Education*, *25*(6), 720–732. https://doi.org/10.1080/13562517.2019.1613352

Owlia, M. S., & Aspinwall, E. M. (1996). A framework for the dimensions of quality in higher education. *Quality Assurance in Education*, *4*(2), 12-20. https://doi.org/10.1108/09684889610116012

Pereira, L., Barreto, R., & Melro, A. (2021). Digital transformation in quality assurance processes: Higher education perspectives. *Education Sciences*, *11*(2), 84. https://doi.org/10.3390/educsci11020084

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Quach, T. T. N., & Nguyen, L. (2024). Using Accreditation Standards for Internal Quality Assurance for Tertiary General English Program in Vietnam. *International Journal of TESOL and Education*, *4*(3), 1-19. DOI: https://doi.org/10.54855/ijte.24431

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). Tertiary Education for the Knowledge Society: Volume 1 and 2. *OECD Publishing*. https://doi.org/10.1787/9789264046535-en

Schindler, L., Puls-Elvidge, S., Welzant, H., & Crawford, L. (2015). Definitions of quality in higher education: A synthesis of the literature. *Higher Learning Research*

Communications, *5*(3), 3-13. https://doi.org/10.18870/hlrc.v5i3.244

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Smith, J., & Naylor, R. A. (2001). Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(2), 389–405.

Stensaker, B. (2008). Outcomes of quality assurance: A discussion of knowledge, methodology and validity. *Quality in Higher Education*, *14*(1), 3–13. https://doi.org/10.1080/13538320802011561

Sun, Y., & Yang, F. (2021). The effectiveness of mobile-assisted language learning in higher education: A meta-analysis. *Educational Technology Research and Development*, *69*(2), 1–20. https://doi.org/10.1007/s11423-020-09821-3

Tam, M. (2001). Measuring quality and performance in higher education. *Quality in Higher Education*, *7*(1), 47-54. https://doi.org/10.1080/13538320120045076

Temple, P. (2008). Learning spaces in higher education: An under-researched topic. *London Review of Education*, *6*(3), 229–241. https://doi.org/10.1080/14748460802489363

Tran, T. M. L. (2024). Blended Learning in EFL Classrooms at a Vietnamese University from Students' Perspectives. *International Journal of TESOL & Education*, *4*(2), 99-117. DOI: https://doi.org/10.54855/ijte.24426

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.

Van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality assessment in higher education. *Higher Education*, *28*(3), 355–371. https://doi.org/10.1007/BF01383995

Yastıbaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessment. *Journal of Language and Linguistic Studies, 14*(1), 178-193.

Yorke, M., & Longden, B. (2004). *Retention and student success in higher education*. McGraw-Hill Education.

Wang, S. (2017). Flipped classrooms in teaching English as a foreign language: A case study in a university setting. *The Journal of Asia TEFL*, *14*(4), 655–668. https://doi.org/10.18823/asiatefl.2017.14.4.5.655

Wang, J., & Cheng, Y. (2018). The impact of quality assurance on student learning outcomes: Evidence from China. *Studies in Higher Education*, *43*(4), 634-649. https://doi.org/10.1080/03075079.2016.1184238

Wiggins, G., & McTighe, J. (2005). *Understanding by design*, expanded (2nd ed.). Alexandria VA: Association for Supervision and Curriculum Development.

Zhang, Y., & Yin, H. (2019). The impact of general English programs on student outcomes: A meta-analytic review. *Journal of English for Academic Purposes*, 38, 1–12. https://doi.org/10.1016/j.jeap.2019.06.001

Biodata

Quach Thi To Nu, a seasoned lecturer at Industrial University of Ho Chi Minh City is currently pursuing her PhD at Ho Chi Minh City University of Social Sciences and Humanities, under Vietnam National University. Her research focuses on improving English teaching methodologies and promoting quality assurance in education.

Prof. Dr. Nguyen Loc is an expert in education management. He has written numerous book chapters published by Springer, contributed to higher education textbooks, and published articles in various educational journals. Additionally, he has been actively involved in SEMEO Retract, language policy development, and the Vietnamese National Foreign Language Projects.

Complexity, Fluency, and Accuracy in Written Works of Vietnamese Learners of English: A Replication Study

Dinh Thao Nguyen[1*]

[1] Faculty of Foreign Languages, Dong Nai University, Dong Nai, Vietnam
[*]Corresponding author's email: nguyendinh.wm18@gmail.com
[*] https://orcid.org/0009-0005-0125-607X
https://doi.org/10.54855/ijte.25522

Received: 19/06/2024 Revision: 26/03/2025 Accepted: 31/03/2025 Online: 23/04/2025

ABSTRACT

Recent studies applying Dynamic Systems Theory (DST) have shown language acquisition's non-linear, indiscrete, chaotic, and highly context-dependent nature. However, limited research has explored these dynamics in English as a Foreign Language (EFL) context. To evaluate DST's applicability in EFL settings, this study partially replicates Larsen-Freeman's (2006) investigation by tracking the language development of four upper-intermediate Vietnamese EFL learners. The study employed the Dynamic Description approach, in which the participants were required to take written tests every three weeks over three months. Their performance across four dimensions - accuracy, fluency, lexical complexity, and syntactic complexity- was examined through both quantitative and qualitative analyses. Although the findings revealed similarities with the original study, the learners' development trajectories differed, proving that the learners were interacting with their surrounding EFL context. These findings underscore the significance of learning contexts on language acquisition and offer valuable insights into more context-sensitive teaching practices.

Keywords: Dynamic System Theory (DST), environment interaction, time-series design, language assessment

Introduction

Traditionally, Second Language Acquisition (SLA) has been viewed as a stage-like, discrete, and linear process, mainly cognitive in nature (Larsen-Freeman, 2006). Influential theories, such as Krashen's *Natural Order Hypothesis* (1982) and McLaughlin et al.'s *Information Processing model* (1983), suggest that language learning progress traverses through separate stages in a consistent manner. However, Dynamic Systems Theory (DST) challenges these notions. Research grounded in the DST framework reveals that learners' linguistic growth is not consistent but rather intertwined and dynamic, shaped by continuous interaction among various elements such as learners' environments, personal motivations, and access to linguistic input

CITATION | Dinh, T. N. (2025). Complexity, Fluency, and Accuracy in Written Works of Vietnamese Learners of English: A Replication Study. *International Journal of TESOL & Education*, *5*(2), 29-56. DOI: https://doi.org/10.54855/ijte.25522

(Larsen-Freeman, 2006; de Bot et al., 2007; Verspoor et al., 2008; Caspi, 2010; de Bot & Larsen-Freeman, 2011).

However, much of the research in this area, including Larsen-Freeman's (2006) – one of the pioneering studies – was primarily conducted in an English as a Second Language (ESL) context, where learners have greater exposure to the target language in their daily lives. In contrast, English as a Foreign Language (EFL) learners typically encounter English primarily in formal instructional settings, making their learning experience distinct. Given that language acquisition is significantly influenced by learners' contexts, objectives, and available input, it is essential to question whether DST findings from an ESL environment can be generalized to EFL learners.

This research explores the reliability and generalisability of Larsen-Freeman's (2006) findings within the Vietnamese EFL context by observing language development in the written works of four upper-intermediate learners over a three-month period. By closely comparing the outcomes of the two studies, the present study will highlight both similarities and discrepancies, thereby providing valuable insights into the dynamics of language acquisition in an EFL environment.

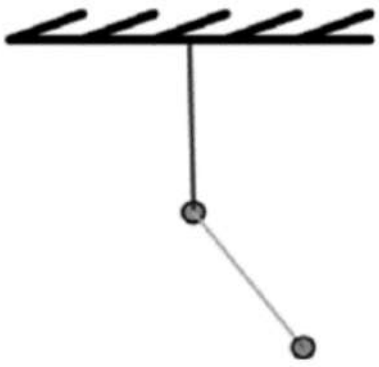
Literature review

Definition of Dynamic System

A system can be defined as a collection of elements that function collaboratively. Each system encompasses multiple sub-systems that interact dynamically with each other (Larsen-Freeman, 1997). Besides, the term *dynamic* in DST refers to the ongoing self-organization driven by internal forces and the surrounding environment (de Bot et al., 2013). Generally, a system is considered a complex or dynamic system - these terms are now used interchangeably - if it demonstrates at least three main features. That is, the system (1) comprises a minimum of two sub-systems that are (2) interlinked with each other but which also (3) demonstrate spontaneous, independent self-organization over time. The movement of a double pendulum – the simplest dynamic system – perfectly exemplifies this concept. The double pendulum consists of two sub-systems: the upper and lower limbs (see Figure 1). Movement in the upper limb causes the lower limb to move chaotically, affecting the upper limb's initial motion, leading to a wholly distinct movement for the entire system.

Figure 1.

A double pendulum

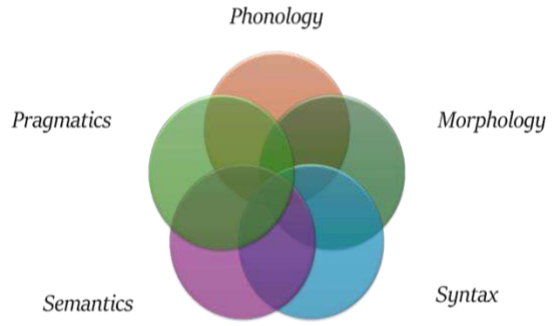


Language as a dynamic system

Larsen-Freeman (1997) argues that language can be explicitly or implicitly perceived as a complex, dynamic system because it exhibits features that are similar to those of dynamic systems. These characteristics include (1) a variety of sub-systems, (2) complete interconnectedness among sub-systems, and (3) autonomous reorganization. To be more precise, language consists of five sub-systems: phonology, morphology, syntax, semantics, and pragmatics, all of which are interdependent (Verspoor et al., 2011) (see Figure 2). For language to function effectively, learners must be able to pronounce words (phonology), associate meanings with those words (morphology), comprehend the meanings of words when combined into sentences (syntax), and grasp how broader non-verbal contexts influence language meaning (pragmatics). Each sub-system varies at different levels depending on internal and external factors (Plotkin, 2006). In summary, DST conceptualizes language as a multidimensional construct that continually reorganizes due to internal changes and contextual adaptations (Morrison, 2006). Consequently, Second Language Acquisition (SLA) is also regarded as a dynamic process (de Bot, 2008).

Figure 2.

The interconnectedness of linguistic sub-systems



Regardless of being a relatively new concept in SLA, DST has quickly become a leading approach for understanding the intricacies of language development in learners by integrating various aspects of SLA. For instance, Verspoor et al. (2008) tracked individual learners' written development and found that progress across accuracy, fluency, and complexity did not follow a steady trajectory but instead displayed irregular patterns driven by personal learning experiences and external conditions. Huynh (2021) also applied DST to explore foreign language learners' anxiety in online language classrooms, illustrating the dynamism and emergence of factors affecting language learning. Furthermore, by transitioning from linear, straightforward cause-and-effect models to non-linear, holistic frameworks, DST offers robust explanations for complex linguistic phenomena such as the "butterfly effect", explaining why learners exhibit significantly different progress even with similar language input.

Second Language Acquisition in Dynamic System View

In the lens of DST, SLA possesses certain characteristics, which I selectively present below.

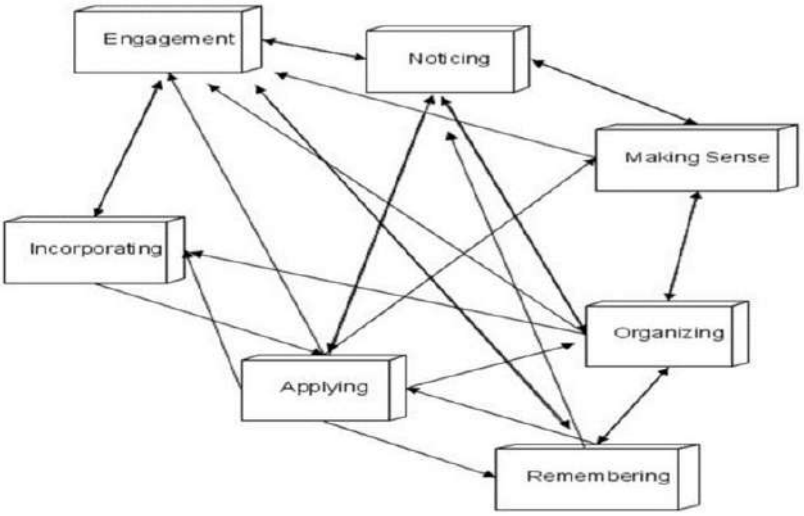
The dynamic system exhibits chaotic variation.

As a result of chaotic variation among and within variables, it is impossible to separate the trajectory of complex systems (Verspoor et al., 2011); thus, learners' performance does not adhere to distinct stages. In the DST-based approach, the information processing (see Figure 3) is a complex process, as each stage could chaotically interact with any or all of the other stages. Sometimes, this complexity may involve iterative procedures or breakthroughs among stages (Harshbarger, 2007). For example, the acquisition of a new word does not guarantee its

immediate integration into a learner's cognition; there may be instances of forgetting, necessitating reassignment of meaning. Alternatively, the application of the new word can influence its meaning, either confirming or contradicting the learner's initial understanding. In cases of contradiction, the learner may revise their understanding or temporarily set it aside until further clarification or engagement occurs.

Figure 3.

A dynamic model of information processing (Harshbarger, 2007)

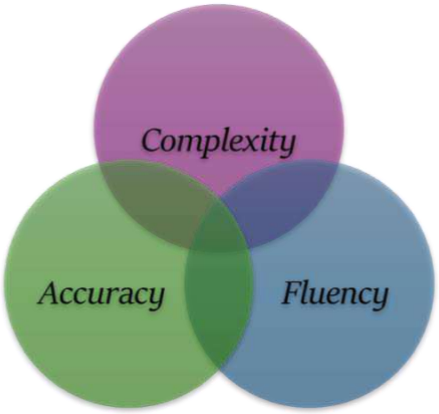


DST is completely interconnected.

In DST, all parts are interconnected, and this applies to language acquisition as well. Phonological, lexical, and syntactic subsystems do not develop separately but are highly interconnected throughout development (see Figure 4); hence, changes in one subsystem influence all others in various ways. An empirical study by Yang and Sun (2015) supports this, revealing that complexity, fluency, and accuracy in language learning are highly interdependent, with progress in one dimension often accompanied by stagnation or regression in another.

Figure 4.

The interconnectedness of linguistic dimensions



However, language sub-systems and dimensions can interact both competitively and supportively, constrained by limited resources such as time, memory, and attention (Robinson & Mervis, 1998; van Geert & Steenbeek, 2008). In other words, there is constant resource competition among sub-systems and dimensions. For example, improvements in complexity might impede accuracy, as seen in Polat and Kim's (2013) study, where advanced Turkish learners displayed gains in lexical and syntactic complexity but no parallel improvement in accuracy. Similarly, Spoelman and Verspoor (2010) observed a regression in accuracy despite improvements in complexity for Dutch learners.

These findings carry significant implications for language teaching and learning. Given the highly interconnected nature of language, assessing learners' progress solely through a single sub-system is inappropriate because it fails to account for the multifaceted interactions of these systems and dimensions over time (Beckner et al., 2009; Larsen-Freeman, 2006, 2009; Perone & Simmering, 2017). Besides, teachers should anticipate potential trade-offs between different dimensions of language development and tailor instruction that strategically balances these competing demands. By doing so, they can create learning environments that support students in effectively managing these challenges and promoting holistic language growth.

The environment is indispensable.

In the DST view, systems experience changes due to the interconnection among sub-systems and their active responses to the surrounding environments (Lewontin, 2000). In the context of SLA, this implies that learners' learning progress is profoundly influenced by their environment (Verspoor et al., 2008). Provided that a language classroom is an environment, learners' performance will be influenced by various contextual factors such as their teachers' voice, classmates' behavior, space of class, and so forth. In simpler terms, SLA arises from an interplay between cognitive development and environmental factors such as the social background of learners, friendships, experiences, and so forth. All these elements continuously interact to shape language in an individual. In fact, numerous studies have examined the environmental influences on SLA. For instance, Lightbown and Spada (2006) show that language enhancement is greatly influenced by the quality of interaction with teachers, underscoring the significance of a supportive learning environment. Similarly, Kinginger (2008) concludes that learners' engagement with peers and active involvement in social activities outside the classroom substantially contribute to language improvements. Another research that clearly indicates the interplay between cognitive development and environmental factors is that of Norton (2000) on immigrant women learning English in Canada. His research proves social identity and the availability of interactional opportunities within the community can remarkably affect language learning outcomes. In short, from a DST perspective, language cannot be exclusively considered as either a cognitive or sociocultural phenomenon, as neither approach alone fully explains the complexities of language learning.

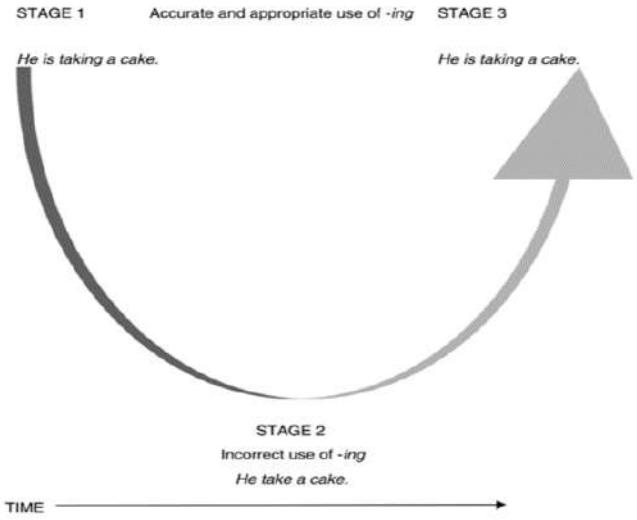
Dynamic systems develop non-linearly.

The dynamic interplay among a system's internal sub-systems and its surrounding environment contributes to another important feature of complex dynamic systems – nonlinearity in development. Due to the presence of multiple interconnected variables, pinpointing cause and effect within the system becomes unfeasible (Godwin-Jones, 2018). In language learning, learners' enhancement does not follow a linear trajectory where one would sequentially move from one piece of knowledge to another (Kramsch, 2012). Instead, when a learner acquires new

information, this knowledge does not simply integrate into the existing knowledge but disrupts the previous state and restructures the entire system (Verspoor et al., 2011). On account of this complex interaction, SLA does not proceed linearly (Murakami, 2016) but is marked by periods of advancement and setbacks (Larsen-Freeman, 1997). The U-shaped development of learners learning grammatical morphemes introduced by Lightbown (1983) (Figure 5) is a typical image demonstrating this learning progress. In the model illustrating the utilization of the continuous morpheme *-ing*, accuracy is initially high but subsequently undergoes a temporary regression before becoming high again.

Figure 5.

T*he U-shaped development: an example of grammatical morpheme –ing (Lightbown, 1983)*



Other researchers have recorded this non-linear pattern in their studies. Nguyen and Pham (2019) found that Vietnamese EFL learners' lexical acquisition exhibited great variability, with periods of impressive progress followed by stagnation or even regression, especially when influenced by external factors, namely academic pressure or limited exposure to authentic language experiences.

There are often multiple routes that are possible among components in dynamic systems.

Complex systems frequently exhibit unpredictable and surprising behaviors among individuals because they consist of numerous dynamic sub-systems (Dörnyei & Murphey, 2003). Furthermore, these sub-systems differ from one another as a result of the complex variations within and around them. A common illustration of this concept is a school of fish. While the group as a whole follows a specific path, each fish swims in its own unique way when observed individually. In terms of SLA, it has been noted that individual differences in motivation, aptitude, cognition, first language (L1), and other factors lead to distinct behaviors, even though the overall group may follow a similar trajectory (de Bot et al., 2007). This intraindividual variability is evident in Larsen-Freeman's (2006) study, which analyzed data from five Chinese participants. She found that although all learners improved in a grand sweep view, each pursued a distinctive developmental pathway.

Research Gap

Despite the fact that Dynamic Systems Theory principles have been studied extensively in language learning, much of the existing research is primarily theoretical, with few empirical studies (Dörnyei et al., 2014). Even when data-driven investigations adopt the DST lens, they often isolate individual characteristics of the theory rather than embracing its holistic view of language learning as a complex, dynamic process (Fischer et al., 2005). This limitation has stimulated demands for more comprehensive emergentism research that studies DST in its entirety rather than in parts. The most thorough analysis of SLA from a complex dynamic perspective is likely provided by Larsen-Freeman (2006), though she is not alone in this endeavor (see de Bot et al., 2005, for example).

However, wholesale studies like Larsen-Freeman's have not been widely applied to the EFL context, where learning environments, interaction nature, surrounding environments, practices, cultural aspects, and so on differ significantly from the ESL context. For instance, unlike ESL learners, Vietnamese EFL learners primarily study English in formal classrooms with minimal exposure to authentic language use outside of school. Moreover, the education system in Vietnam heavily focuses on grammar and reading comprehension, often at the expense of speaking and listening skills. This is further compounded by a curriculum that emphasizes rote learning and test preparation. Thelen and Smith (2006) suggested that the environment is a crucial factor in language acquisition, raising the question of whether Larsen-Freeman's findings apply to both ESL and EFL contexts. The dearth of replication studies to validate existing findings remains a significant gap. These reasons have encouraged me to carry out a replication study in the EFL context (Vietnam), in which I closely mirror Larsen-Freeman's data collection and methodology to assess the reliability and generalisability of her findings.

Research Question

The research aims to address the question below from the aforementioned research gaps and objectives.

To what extent can Larsen-Freeman’s (2006) findings be applied in the EFL context?

Methods

Methodology & Participants

When investigating second language acquisition from a dynamic perspective, researchers must focus on three key elements: *time, complexity, and interaction with the environment*. Therefore, I concur with Larsen-Freeman and Cameron (2007) that traditional methodologies that merely capture single points in time are unsuitable. Instead, new approaches are required to evaluate the entire spectrum of learners' linguistic repertoires over time and allow environmental factors to be embodied. Two highly recommended methods to study dynamic systems include Retrodictive Qualitative Modelling (RQM), introduced by Dornyei (2014), and Dynamic Description, proposed by van Gelder and Port (1998). The former approach addresses DST's limited predictability by identifying outcomes and then retrospectively exploring developmental pathways. However, since the focus here is on the learning process rather than outcomes, I decided to opt for The Dynamic Description since it has been proven to be "a general conceptual apparatus for understanding the way systems – including, in particular, non-linear systems – change over time" (van Gelder & Port, 1998, p.17).

This study adopted a longitudinal, time-series design to provide a dynamic description of learner development. As a partial replication of Larsen-Freeman's study, it closely followed her methodology regarding sample size and participants' proficiency. However, unlike the original research, which took place in the United States - an ESL environment, this study was conducted in Vietnam, offering insights into language development in an EFL context. Five upper-intermediate Vietnamese TESOL teachers, aged between 23 and 30, were chosen. They are confidentially referred to by the letters D, T, G, M, and N. T, G, M, and N work part-time in language centers, while D is a full-time high school teacher. Their high level of English proficiency is evidenced by their IELTS certificates, with overall bands ranging from 6.0 to 7.0.

In regard to instruction, the participants took part in a 2.5-hour class on a weekly basis over three months. A communicative book primarily focusing on grammar was used according to the participants' needs. Between classes, they were assigned daily self-study tasks designed to replicate those of the original study. These tasks included textbook homework, 5 minutes devoted to pronunciation practice (i.e., listening to the radio and paying attention to specific words or sounds pronounced differently from how the participants might say), 15 minutes of free reading, and one hour of listening or watching TV to identify new words. The researcher did not grade or regularly monitor the self-study tasks to observe how participants autonomously organized their L2 learning systems. Additionally, it is noteworthy that participants T, G, and M were further participating in IELTS preparation courses for an upcoming re-examination alongside my class.

Regrettably, participant N withdrew from the study after one month and a half, which not only resulted in an interruption to the data collection process but also lost the faithfulness of the sample size to the original study. However, recruiting a replacement was deemed impractical due to time constraints and concerns about introducing systematic differences. Instead, the study proceeded with the remaining participants to maintain research validity and ensure completion within the allotted time frame. Finally, the study has four participants named T, G, D and M.

Data collection

Over three months, once every three weeks, the participants were asked to accomplish the same task: writing narrative stories about a past event without a dictionary consultation. The choice of topic and the length of the narratives were entirely at their discretion, aiming to create natural conditions to accurately assess the learners' development across all linguistic domains. The assessments were untimed, and I did not receive feedback. All participants undertook the initial test after three weeks of instruction. This baseline data diverges from Larsen-Freeman's research, where the initial test was conducted after four months. The modification was made to maintain a comparable number of tests within the study's timeframe.

Data analysis

The four learners' writing was analyzed through quantitative and qualitative methods. This dual data analysis approach helps provide a comprehensive understanding of the changes in learners' language abilities.

Quantitative analysis

In terms of quantitative analysis, the data were analyzed at both macro- and micro-levels. On the macro scale, an overall picture of language improvements in fluency, accuracy, grammatical

complexity, and lexical complexity across four participants was obtained by analyzing their written works which had been segmented into "t-units" - “one main clause with all subordinate clauses attached to it" (Hunt, 1965, p.20). A summary of the language dimensions and the associated measures is provided in Table 1. Subsequently, the mean values for each dimension are graphed on line charts to depict the group's progress over the three-month learning period.

Table 1.

The language dimensions and their measures

Language Dimensions	Measures
Fluency	average number of words per t-unit
Accuracy	the proportion of error-free t-units to t-units
Syntactic Complexity	average number of clauses per t-unit
Lexical Complexity	type-token ratio (the ratio of different words to total words)

On the contrary, the micro level details how each participant's developmental path evolved over time by plotting the four language indices on separate line graphs. Furthermore, to accentuate the intraindividual variations among the learners and facilitate cross-dimensional comparisons, performance measures were transformed into z-scores and represented graphically. This approach made it possible to observe distinct developmental patterns for each learner across different language indices.

Finally, to further consolidate the learners’ intraindividual differences, the rate of change over time was also measured. More specifically, to calculate the differences between consecutive tests, the initial test was set as the baseline (assigned a value of 0), then the value of each subsequent test (e.g., test 2) was divided by that of the preceding one, and the number received was the rate of change.

Qualitative analysis

In this analysis, the written products of four participants were broken into idea units—"a message segment consisting of a topic and comment that is separated from contiguous units syntactically or intonationally" (Ellis & Barkhuizen, 2005, p.154). These idea units were then juxtaposed in a table for each participant to observe their narrative construction and change over time.

Research Findings

Quantitative analysis

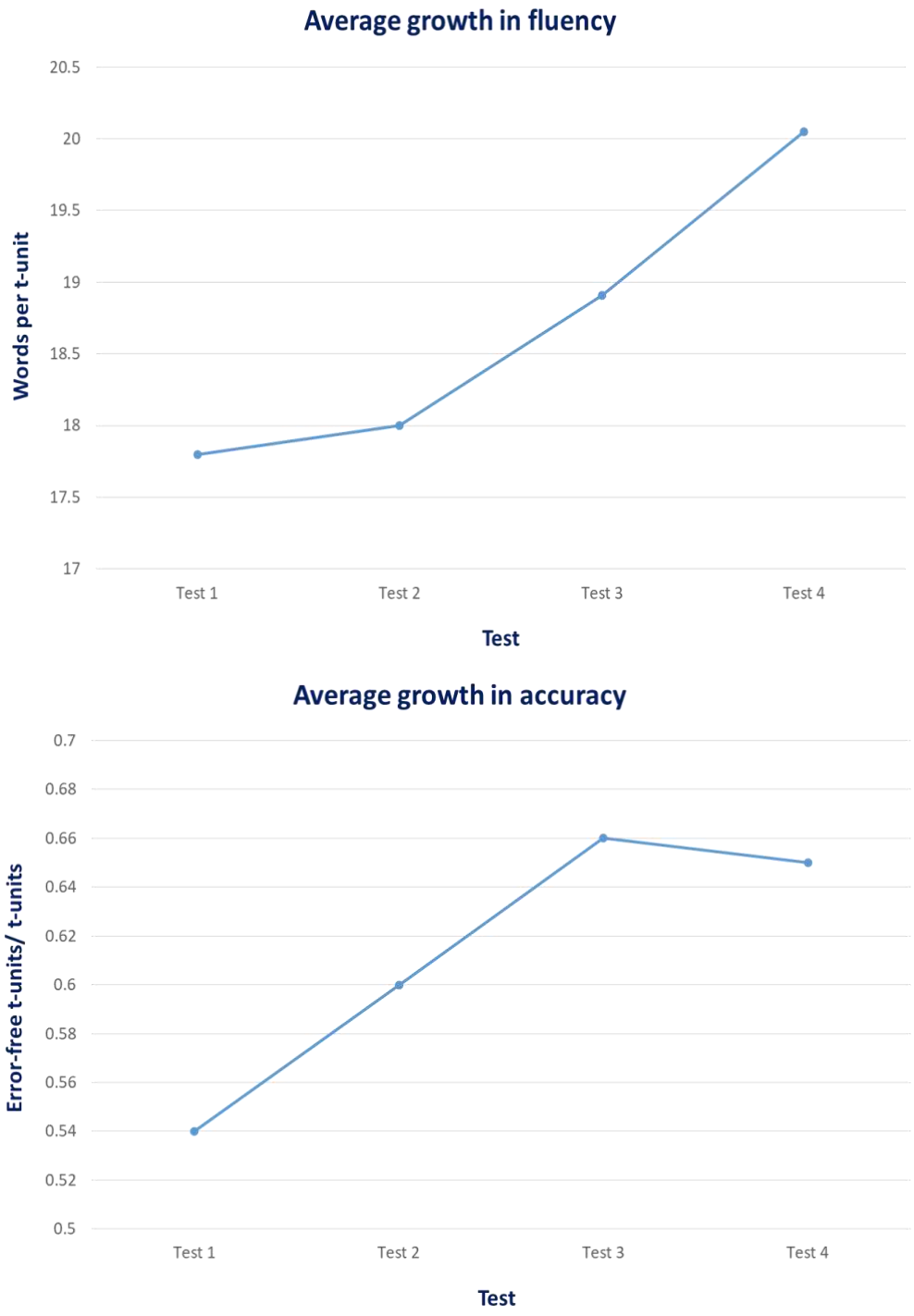
Group Development

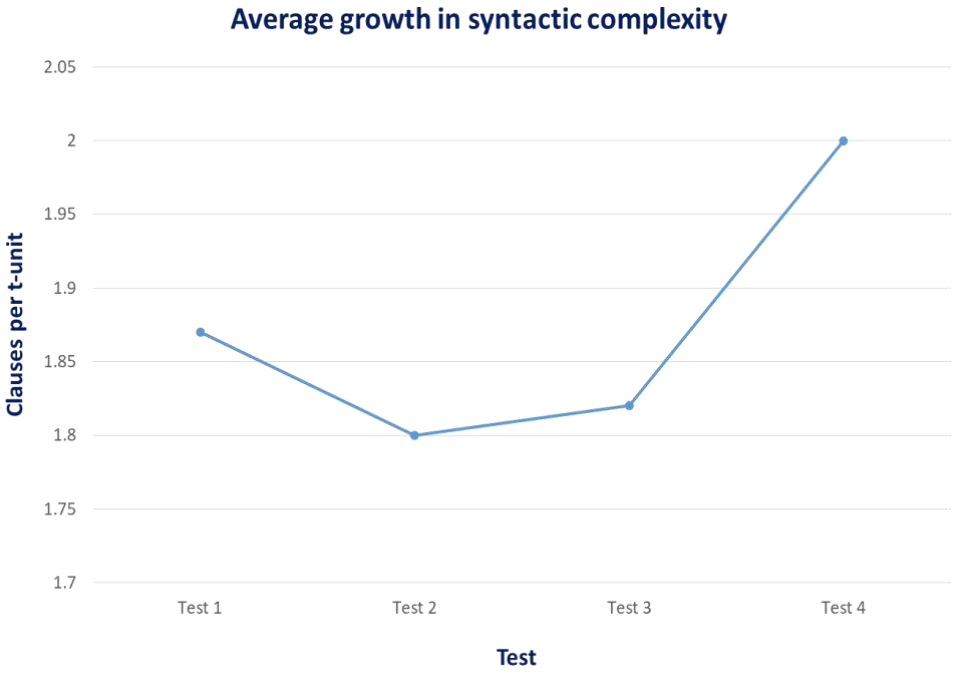
The mean scores across four dimensions of language proficiency - fluency, accuracy, lexical complexity, and syntactic complexity (see Figure 6) - showed overall progress among the participants. However, the development patterns varied across these indices. Fluency showed significant growth, while accuracy initially improved but marginally decreased in the final test.

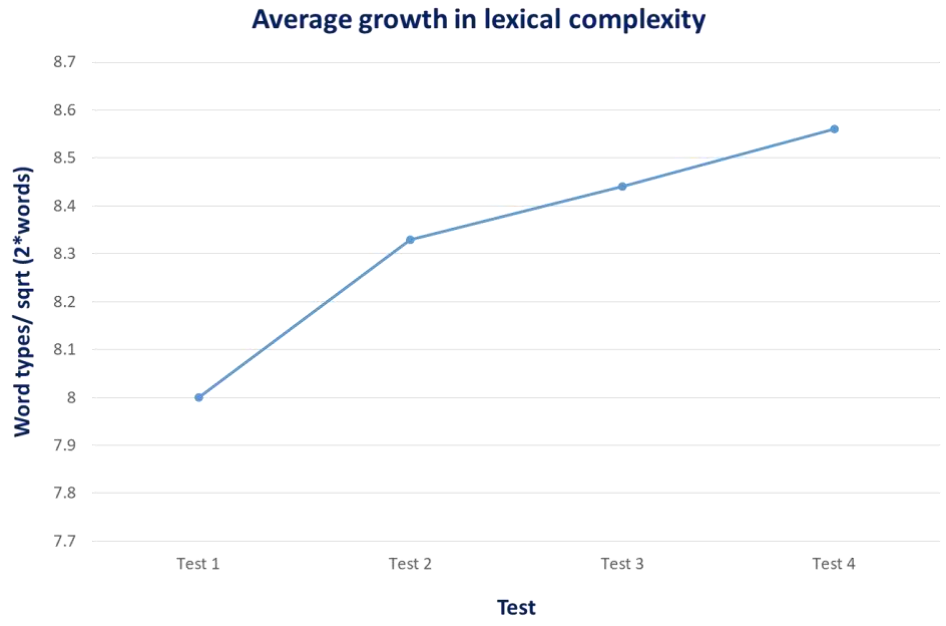
Syntactic complexity witnessed some fluctuations over time but achieved certain levels of achievement in general. Finally, lexical complexity within the group demonstrated stable progress. Overall, compared to three months ago, participants' writing exhibited greater fluency, accuracy, and complexity in both grammar and vocabulary.

Figure 6.

Group averages (± 1 SD) over time on four indices







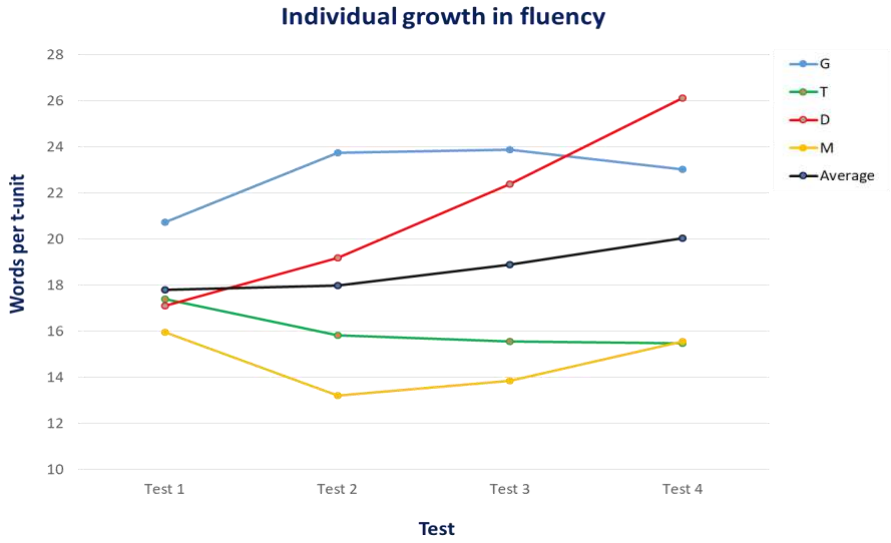
Nonetheless, analyzing data based on group averages solely captures the process or functional relation, leading to a lack of validity for individual variability (Sidman, 1960). Therefore, disaggregating the data would unveil a different perspective on learners' development.

Interindividual variability

When each index is charted individually, a more intricate picture of the developmental paths across the four indices emerges. All line graphs from Figures 7 to 10 describe individual growth patterns across accuracy, fluency, lexical complexity, and syntactic complexity.

Figure 7.

Interindividual growth in fluency over time and the average for four participants

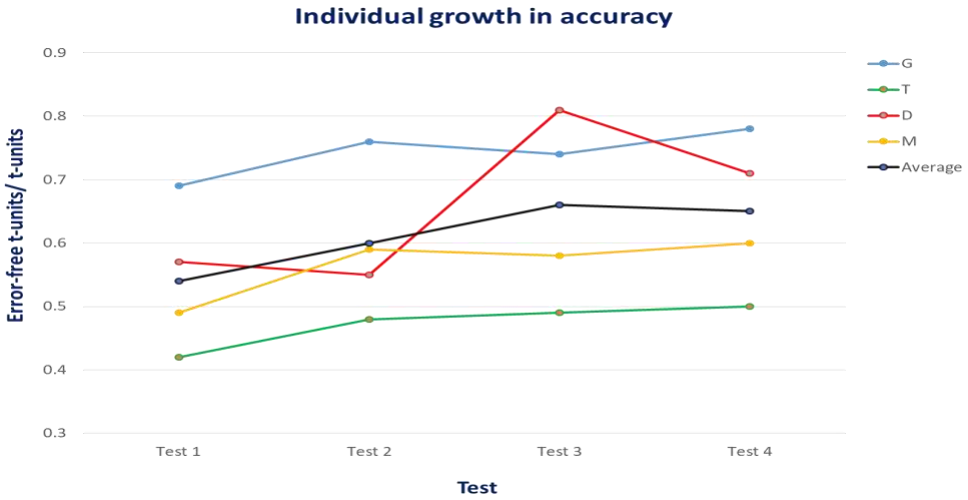


While group averages depicted a relatively smooth upward trend, individual performances were characterized by peaks and valleys. Some exhibited progress, others experienced setbacks, and some remained unchanged after three months. For instance, individual fluency growth indicates that D made significant gains; T's performance gradually declined; while G and M oscillated between advancement and regression.

With regard to accuracy (Figure 8), G, T, and M saw modest variations over the four tests. Although they showed some improvements in accuracy by the end of the study, the changes were unremarkable. Conversely, D's accuracy enhancement fluctuated drastically. Her accuracy decreased modestly in test 2 but surged in test 3 and then declined again in the last test.

Figure 8.

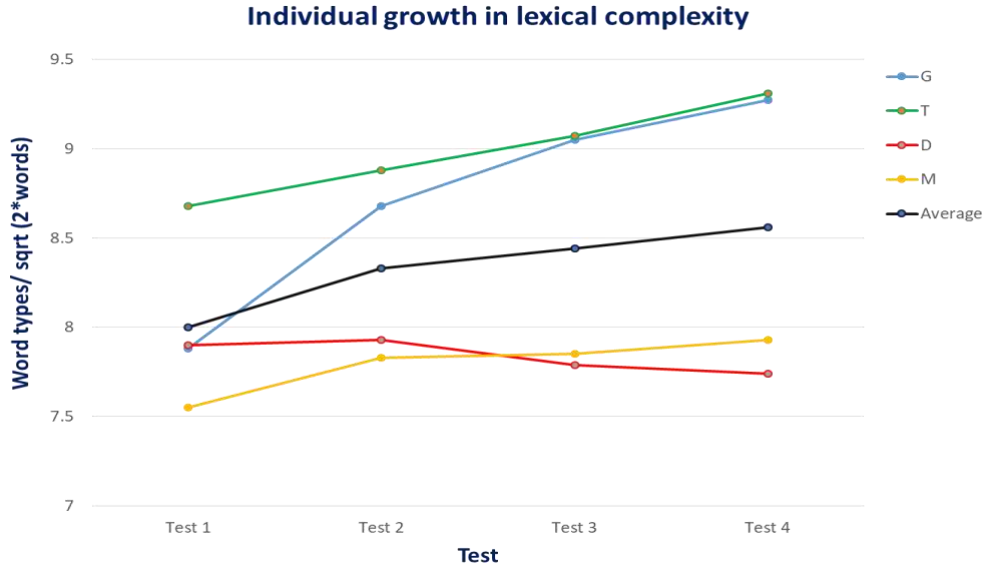
Interindividual growth in accuracy over time and the average for four participants



In terms of lexical complexity (Figure 9), participants G, T, and M consistently expanded their vocabulary sizes throughout the tests, whereas D showed a gradual drop over the same period.

Figure 9.

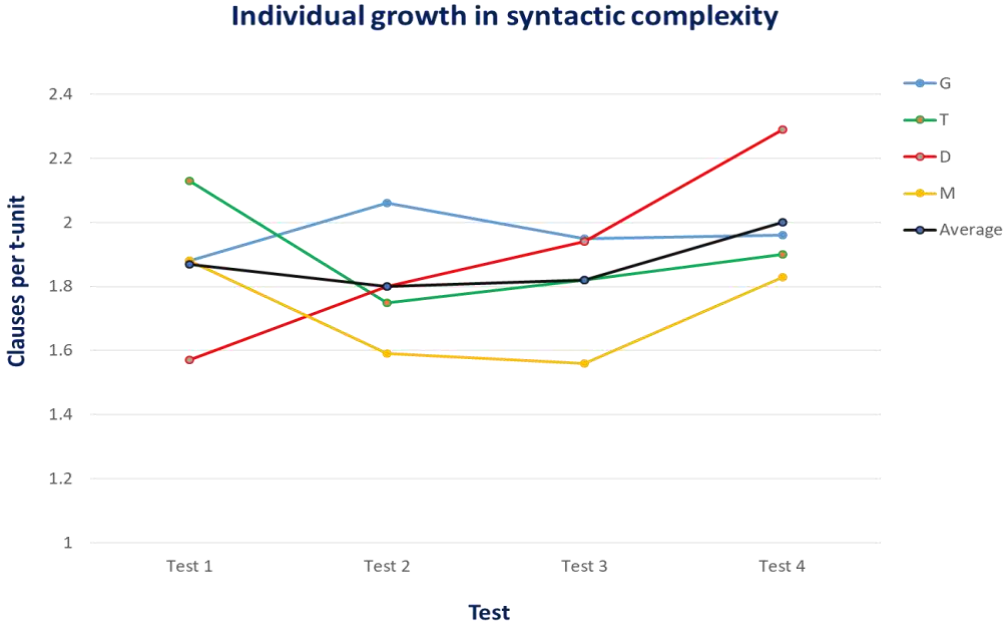
Interindividual growth in lexical complexity over time and the average for four participants



Ultimately, the individual progression in syntactic complexity illustrated in Figure 10 shows that D had a steady upward trajectory, whereas others experienced fluctuating developments. To be more specific, G's grammar advanced in test 2, but T and M did not. However, while G did not show further improvement in subsequent tests, both T and M demonstrated progress.

Figure 10.

Interindividual growth in syntactic complexity over time and the average for four participants



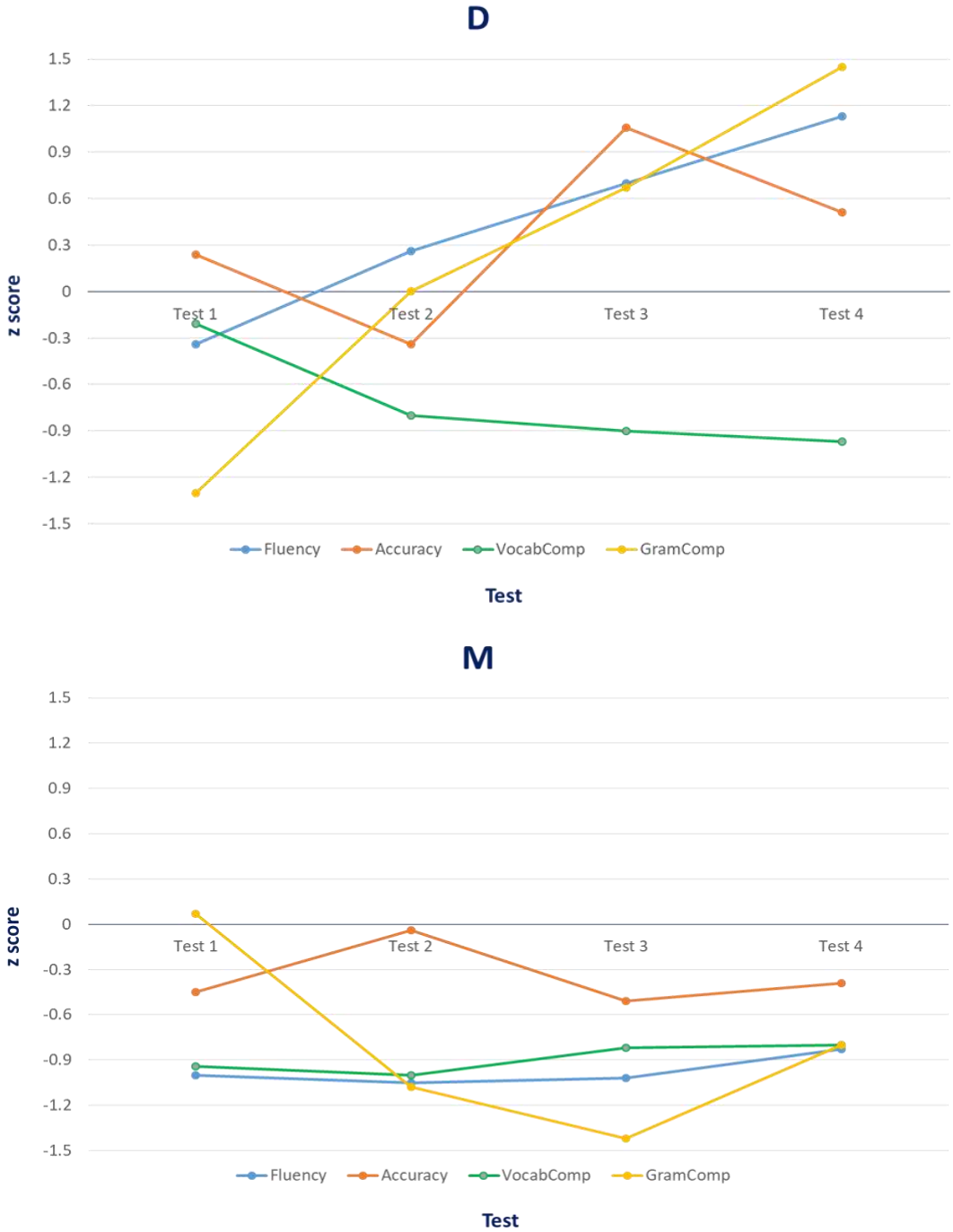
Intraindividual variability

Traditionally, intraindividual variability, which refers to temporary fluctuations among subjects,

has constituted a form to measure errors (Costa et al., 2019). However, from the lens of DST, this within-subject variability serves as a fundamental source of information to illuminate the learners' developmental paths (van Geert & Steenbeek, 2005). To highlight the intraindividual differences between the four learners, all performance measures were converted into z-scores (Figure 11). Since the z-score allows comparability across four language factors, a more accurate portrayal of each learner's progress over time can be captured.

Figure 11.

Intraindividual variation over time for four participants on four indices



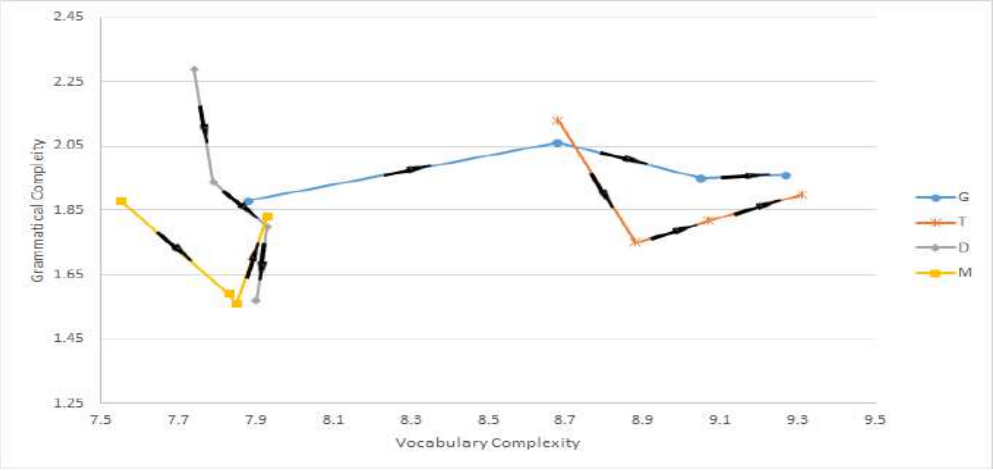
The four line graphs in Figure 11 reveal that each learner's language development pathway is dynamic and complex. Throughout the study time frame, participant G concentrated a great deal on vocabulary, with lexical complexity being the only dimension that progressed. T, on the other hand, experienced a decline across all four indices, particularly in grammar. In contrast, D enhanced in almost every dimension except for vocabulary. Finally, M exhibited no remarkable improvements in any indices. These outcomes proved the learners possess distinct achievement orientations over time.

Two dimensions of the participants' performance - grammatical complexity versus vocabulary

complexity and grammatical complexity versus fluency - are plotted on graphs to obtain more in-depth insights into the individuals' favored developmental paths. The selected indices imitate those used in Larsen-Freeman's (2006) research. Figure 12, comparing the progression of grammar and vocabulary among four participants over the three-month period, shows that G, T, and M made strides in lexical complexity, with G demonstrating particularly notable improvement, while D focused more on enhancing syntactic complexity.

Figure 12.

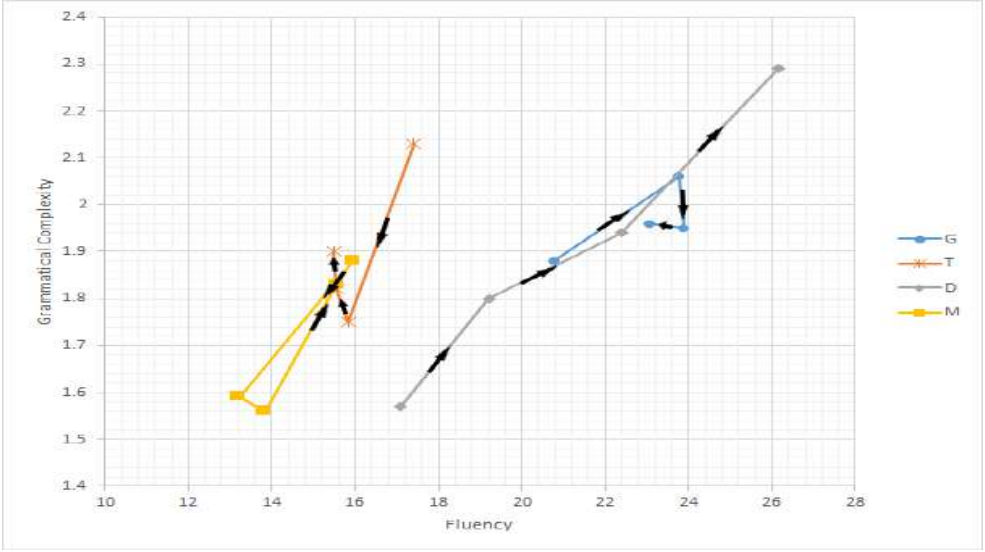
Change of grammatical complexity compared with vocabulary complexity for four participants



Simultaneously, when graphing grammatical complexity against fluency, participant G made significant progress in grammar, whereas the others fell somewhere between these two dimensions (Figure 13).

Figure 13.

Change of grammatical complexity compared with fluency for four participants

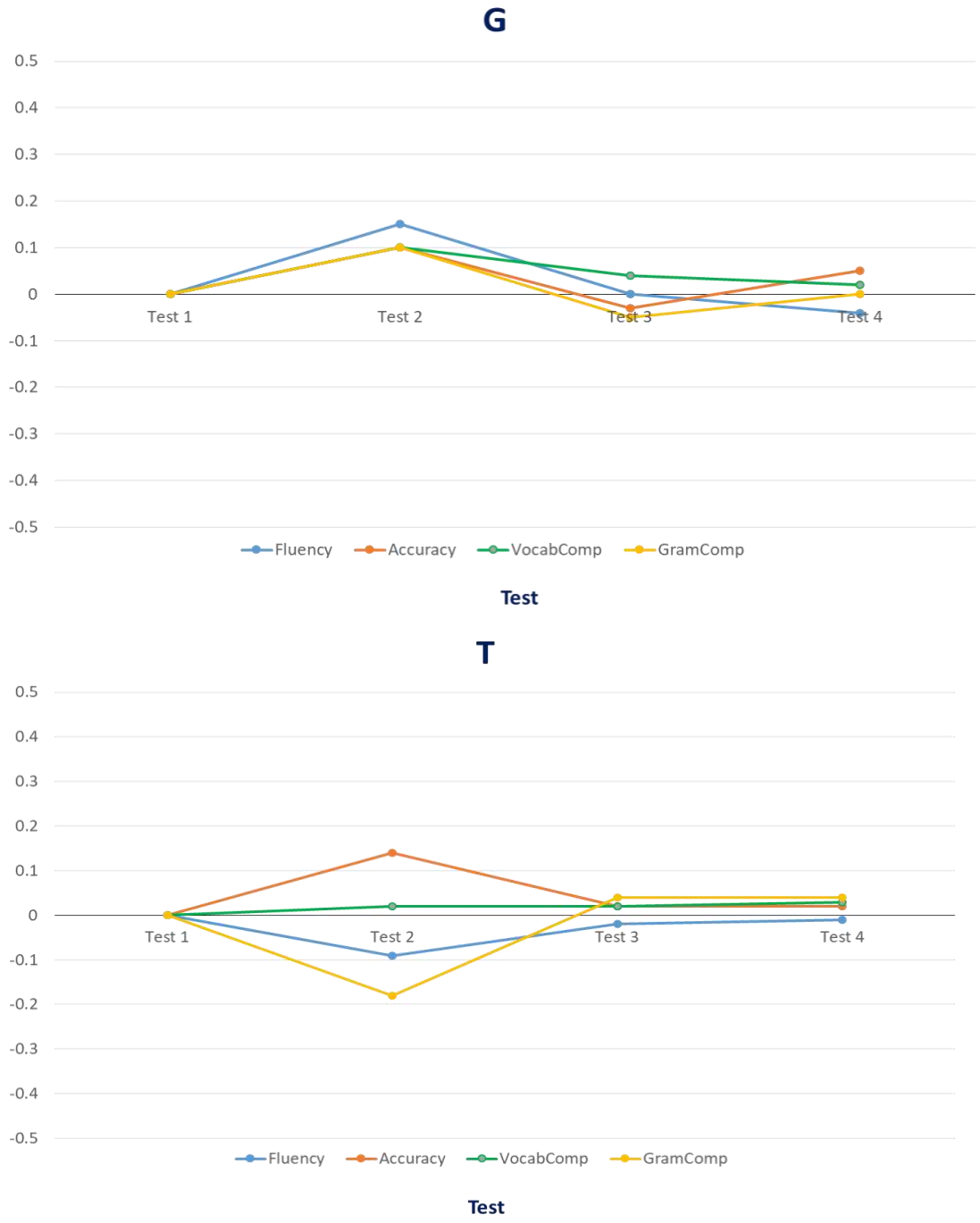


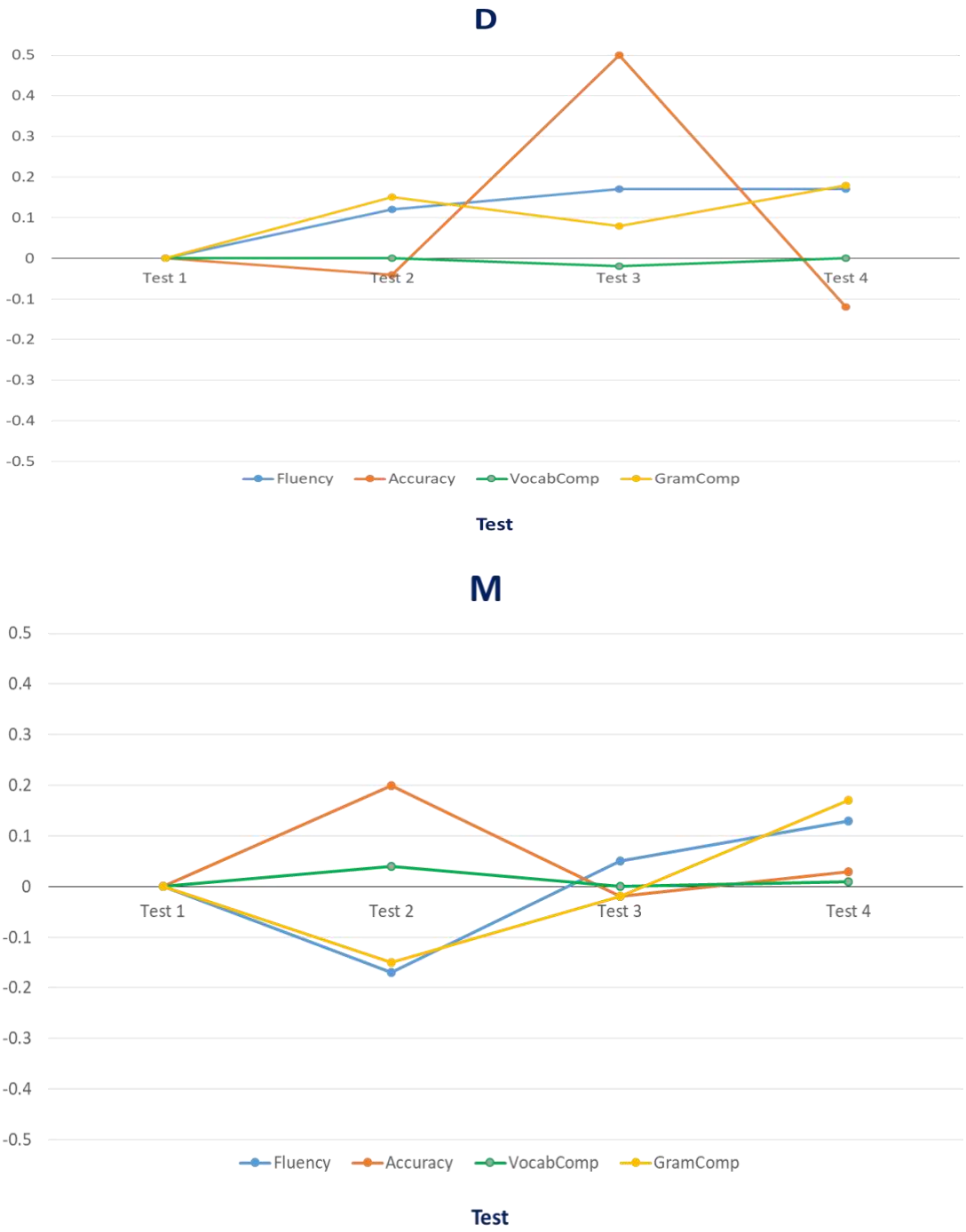
The rate of change fluctuated unevenly among the four participants (see Figure 14), highlighting

the individual differences in language development. To be more precise, both G and T's growths were gradual across all indices, yet T's syntactic complexity rose faster. In contrast, D and M exhibited a swift increase in accuracy but slower in lexical complexity. Overall, the most substantial rate of change occurred in accuracy, whereas vocabulary complexity had the slowest progression.

Figure 14.

Rate of change in four indices for four participants over time





In conclusion, the quantitative analysis shows that the group overall experienced improvements in fluency, accuracy, syntactic complexity, and lexical complexity. However, each participant had a unique developmental trajectory. Furthermore, their progress was marked by fluctuations rather than a steady upward trend, highlighting their language development's chaotic and non-linear nature.

Qualitative analysis

Owing to spatial constraints, in this qualitative analysis, I will selectively present certain data

to highlight specific language performance aspects under the DST view. The chosen data primarily derive from narrative stories provided by participant G, who holds a bachelor's degree in English Language Teaching and has four years of experience as a teacher of English. She works part-time at a language center and also tutors several private classes at home. Table 2 contains seven idea units arranged in a table to facilitate comparison. All of the original features, including grammatical errors, have been preserved.

Table 2.

Participant G's written story data from four tests (seven idea units)

Idea Unit	***Test 1***	***Test 2***	***Test 3***	***Test 4***
1		As soon as the Tet holiday passed, the Vietnamese government gradually implemented the social distancing policy, fearing it would get out of control due to a lack of medical aid and equipment.	As soon as the Tet holiday passed, the Vietnamese government gradually implemented a social distancing policy throughout the country, fearing that it would become out of control due to a lack of medical aid and equipment.	As soon as the Tet holiday passed over, a social distancing policy was gradually conducted throughout the country by the Vietnamese government in fear of being out of control for lacking medical aid and equipment.
2	It was exactly the day after the Tet holiday that my manager organized an urgent meeting and carried out an online teaching workshop, which enabled teachers to help students avoid being interrupted in their learning due to the pandemic.	Immediately, my center manager organized an urgent meeting to figure out some solutions to help students not be interrupted in their learning process. The meeting ended up choosing the Zoom app as an interactive teaching method over the distance.	Immediately, my center manager organized an urgent meeting to devise some online teaching methods that would help students avoid being interrupted in their learning process.	Immediately, my center manager organized an urgent meeting to devise some online teaching methods that would help students avoid being interrupted in their learning process.
3				Obviously, online teaching was altogether supported to be the best solution to such a challenging situation.

4	At that time, the Zoom app was the best choice for nearly all schools and institutions to inspire students with knowledge at a distance.	*(..., which ended up choosing the Zoom app as an interactive teaching method through a far distance.*)	At that time, Zoom became the best choice for nearly every school and institution, which could enable teachers to interact with their students with some simple tools.	At that time, Zoom became the most appropriate application for nearly every school and institution carrying out online classes, which could enable teachers to interact with their students with some simple tools.
5	I spent a complete day preparing for a detailed lesson plan **on** the first trial day.	My extreme insecurity made me spend the whole day preparing a detailed lesson plan **for** the first trial day.	My extreme insecurity made me spend the whole day preparing a detailed lesson plan **for** the first trial day.	**On** the first trial day, while I was struggling to prepare a detailed lesson plan, my insecurity gradually increased in my mind
6				Sometimes, I felt that I could not control my rapid heartbeat.
7	I was so stressed out that I **went to** the center very early to set up my laptop and microphones.	I was so stressed that I **came to** the center very early to set up my laptop and microphones.	I was so stressed that I **came to** the center very early to set up my laptop and microphones.	In fact, I was so stressed out that I **came to** the center very early to set up my laptop and microphones.

Several notable differences are observed in the narrative. One of the most outstanding features is the emergence of new idea units in later tests that were not present in the initial one. For example, idea unit #1 was absent in the first test, and idea units #3 and #6 only appeared in the fourth test. Another significant difference is the increased complexity in viewpoints seen in the last test (e.g. a detailed description of the government's response to the coronavirus pandemic in idea unit #1) and changes in verb tense compared to earlier versions. This complication in tense is obvious in idea unit #5, where G shifted from using the Simple Past tense in tests 1, 2, and 3, *"I spent a complete day preparing..."* to the Past Continuous in test 4, *"...while I was struggling to prepare..."*. The alteration in tense also reflects a change in the narrative focus from describing actions *"preparing for a detailed lesson plan..."* to expressing emotions *"my insecurity gradually increased..."* Given that G is doing a storytelling task with the teacher as a targeted reader, her use of language probably aims to evoke the teacher's empathy by sharing her growing feelings of anxiety.

During the study, the inconsistency in language use, especially the alternation of prepositions, was also observed (e.g., "on" and "for" in idea unit #5). Language instruction might have influenced these shifts in preposition usage by emphasizing distinctions between these prepositions in class. Another common pattern is the ephemeral nature of language performance. This is exemplified in idea unit #7, where the verb "went to" used in test 1 was later replaced by "came to"despite both versions being grammatically correct. Such transient language forms are also recorded in the narrative stories written by Participant D, a high-school teacher (Table 3). That is the verb "tasted" in idea unit #1, initially used incorrectly in the first two tests, was corrected to "taste" in test 3.

Table 3.

Participant D's written story data from four tests (one idea unit)

Idea Unit	***Test 1***	***Test 2***	***Test 3***	***Test 4***
1	For the rest of our travels, we decided to visit the neighborhood islands to learn more about the residents' lives and **taste** their special cuisines.	For the rest of our traveling, we decided to visit the neighborhood islands to learn more about the local culture and **taste** their special cuisines.	For the rest of our travels, we decided to visit the nearby islands to learn more about the local culture, **taste** their special cuisines, and dive to see colorful reefs and underwater creatures.	We tried out so many new things there, such as surfing and diving to see the colorful coral reefs and beautiful underwater creatures.

Discussion

This study aims to assess the reliability and applicability of Larsen-Freeman's (2006) research findings in the EFL (Vietnam) context to construct a more comprehensive understanding of how DST affects language acquisition. The research results revealed both similarities and differences compared to Larsen-Freeman's. As the data were analysed via both quantitative and qualitative methods, the discussion will be structured in these two types of data.

Quantitative Analysis

At the macro level, the average group performance indicates advancements in four indices: fluency, accuracy, lexical complexity, and syntactic complexity. These findings align with Larsen-Freeman and other researchers like Bygate et al. (2001), who argue that repeated tasks improve accuracy and fluency. At the micro level, an analysis of interindividual progress over the period of three months reveals varying trajectories compared to group averages, which is in accord with Larsen-Freeman's. These divergent developmental patterns demonstrate that language is dynamic and chaotic as a complex system.

The results from intraindividual variability show that the principle features of DST are evidently met, namely nonlinearity in development, complete interconnectedness, interaction with the environment, and individual variation. To be more precise, the four participants displayed different performances, characterized by periods of growth and decay. Even within each learner's development, all dimensions had no uniform developmental pathway. This highlights the necessity to differentiate specific components, as Norris and Ortega (2009) suggested, and to adopt more personalized approaches to language assessment, as stated by Polat and Kim (2013). The finding is in agreement with Larsen-Freeman's research, which observed significant variation in performance among five Chinese learners. Eskildsen (2012), investigating the utilization of adult learners' multi-word expressions, also revealed that his participants' progression unfolded unevenly, affected by contextual factors and prior language exposure. Similarly, Dong (2016) highlighted non-linear patterns in EFL learners' writing development, indicating that improvements in a particular language domain sometimes coincided with other regressions. These studies affirm that nonlinearity is a consistent phenomenon in language learning across both ESL and EFL contexts.

The findings also prove that language acquisition operates as a dynamic system with interconnected elements and involves trade-offs among different dimensions. Obviously, there was a resource competition between accuracy and lexical complexity in the linguistic development of the four learners, though this was less noticeable for Participant T. The increased emphasis on accuracy can be attributed to the participants' roles as teachers of English. In Vietnam, English teachers are required to be absolutely precise about language use since they are expected to be models for learners to follow. Thus, they have to maintain high accuracy levels in their language (Lewis, 2002). This expectation will likely influence these participants' learning experiences as the learner's performance adjusts according to different contexts to establish their own developmental path (Larsen-Freeman, 1997). This explanation can also clarify why T experienced less "intense" competition. Among the four participants, T has the least teaching experience and has only recently taught one or two classes. Thus, she is less influenced by the aforementioned belief. A similar observation regarding the interplay between accuracy and complexity was noted in Larsen-Freeman's study, possibly due to the unique circumstances of the learners involved as well. All of her participants were professionals who needed high accuracy in their careers. However, Polat and Kim (2013) found a stark difference in their results with advanced learners in Turkey, who showed significant improvements in complexity but no corresponding development in accuracy after a year. Again, this disparity can be because of the learners' context; immigrant workers learn English through everyday communication, targeting efficient communication over strict accuracy. These outcomes illustrate that language development is not a linear process but rather adaptive and responsive, shaped by the surrounding environment to create its own developmental dynamics.

When plotting lexical complexity against syntactic complexity, it is intriguing to note that most of my learners, except participant D, placed greater emphasis on vocabulary. This contrasts with Larsen-Freeman's findings, where four out of five participants in her study prioritized grammar. Although replicating Larsen-Freeman's exact language instruction was not feasible, her key foci on grammar were faithfully maintained in this research. Hence, the observed discrepancy in language learning outcomes may stem from differences in learners' priorities.

Among the four participants, G, T, and M were concurrently preparing for an IELTS examination and attending additional preparation courses alongside regular classes, unlike D, who worked full-time as a high school teacher, did not. In Vietnam, the IELTS holds vital implications for the life opportunities of test-takers. Nguyen (2025) also indicates that high-stakes tests like IELTS can exert considerable washback effects on students' learning behavior, potentially influencing their learning outcomes and shaping their focus and strategies. Moreover, these IELTS preparation courses typically emphasize various linguistic components, with vocabulary learning being particularly targeted due to its perceived importance for success (Drummond, 2018). Therefore, learners preparing for the IELTS exam may prioritize vocabulary acquisition more than grammar. Conversely, D, as a high school teacher, may give more priority to grammar due to its central role in high school language instruction in Vietnam (Lewis, 2002). Despite the differences in results compared to the original study, via analyzing the participants' context, this discrepancy further highlights the contextual triggers behind developmental divergence.

Qualitative Analysis

The qualitative findings reinforce the dynamic interaction between learners' environments and their evolving linguistic systems. Take Participant G's writing as an example; her reference to

the Vietnamese government's response to the Coronavirus pandemic in test 2 coincided with Vietnam's second national lockdown, evoking reminiscence of her previous experience during the first lockdown shortly after the Tet holiday. Similarly, in a study carried out with 28 EFL learners, Ngo (2025) indicates that learners' cultural backgrounds can positively influence language development. These studies underscore the role of cultural and social engagement in developmental processes. Furthermore, the appearance of new idea units and self-correction attempts (e.g., D's endeavor to self-correct the verb "tasted" to "taste", G's struggles with the prepositions "on" and "for" or between the verbs "went to" and "came to") emphasizes the emergent nature of linguistic development. These phenomena align with Larsen-Freeman's key DST features, yet this study further contextualizes these changes within a culturally distinct educational context.

The study's findings challenge the assumption of linear developmental patterns, demonstrating that learners' unique socio-academic contexts drive distinct linguistic priorities. This research enriches the DST framework with a more context-sensitive perspective by shedding light on how contextual factors like individuals' social positions and exam pressures interact with internal language systems.

In conclusion, this study corroborates the nonlinearity and context-dependence identified in Larsen-Freeman's study while also demonstrating how particular cultural and educational factors in Vietnam actively shape learners' dynamic trajectories. These insights advocate for more individualised assessment approaches and highlight the necessity of accounting for socio-contextual influences when teaching and assessing learners.

Conclusion

This research aimed to validate Larsen-Freeman's (2006) findings in the EFL context, where learning environments differ significantly from ESL settings. A replication study closely mirroring the original study's methodology, including participant proficiency levels, tasks, instructions, data collection, and analysis, was conducted to attain this goal. Through quantitative and qualitative analyses of narrative stories from four Vietnamese upper-intermediate learners, it can be concluded that Larsen-Freeman's findings are applicable in the EFL context. Most of the Dynamic Systems Theory (DST) characteristics identified in the original study were recorded. Despite some discrepancies, these differences support Larsen-Freeman's assertion that the interaction between learners and their environment shapes language development. It can be concluded that Larsen-Freeman's findings are applicable to the EFL context.

Finally, it is worth noting that the sample size of this research is small - only 4 participants. In longitudinal, DST-based research like mine, the goal is to uncover patterns of individual variability over time rather than broad generalizations. This sample size allows for in-depth insights into how each learner's trajectory unfolds in response to internal and external factors, capturing language development's dynamic and context-sensitive nature. However, the targeted participants proficiency in this study are upper-intermediate learners who demonstrated a strong motivation for learning English. Since learners at varying proficiency levels may exhibit different developmental patterns (Larsen-Freeman, 2009), these findings should not be generalized to lower-proficiency learners. Therefore, future research could expand the participant pool across proficiency levels — from beginner to advanced — to provide broader

populations and more comprehensive analysis.

Implications

Regarding pedagogical implications, the study's findings underscore the importance of task repetition for language enhancement. Therefore, language teachers should ensure learners have ample opportunities to repeatedly revisit the same or similar content. For instance, learners could read a passage accompanied by listening or viewing materials on the same topic. Moreover, by having students engage in repetitive tasks over time, teachers can observe and identify changes in learners' language skills, including subtle nuances. This enables teachers to determine when and what type of scaffolding is necessary to support learners effectively.

Moreover, as language constitutes a complex interconnected structure encompassing multiple dimensions, educators should refrain from isolating specific elements when assessing a learner's progress. Rather, they should consider other factors such as the interactions among these competencies and, most crucially, the learner's current circumstances.

Ultimately, the diverse developmental paths of each learner highlight the significance of individual differences in the learning process. Therefore, instead of enforcing rigid, standardized, one-size-fits-all materials and pedagogy, educators and curriculum designers should cultivate an input-rich learning environment that integrates various engaging activities to foster individualized learning experiences. Utilizing online resources can assist teachers in sourcing materials tailored to students' needs and interests. Moreover, recognizing that each learner may possess unique achievement orientations, language courses should prioritize independent learning in conjunction with stated learning objectives for each lesson.

Apart from the educational implications discussed earlier, there are some methodological implications. Since DST inherently focuses on changes over time, time itself is a crucial element in DST-based studies. Therefore, conducting a longitudinal study becomes essential to capture meaningful time-series data. However, adjustments in the frequency and intensity of observations may be necessary to align with the desired timeframe.

Another important implication of the research is the use of VocabProfile to gauge lexical complexity. Given its efficiency and ease in identifying types and tokens, future researchers should consider integrating this tool into their analyses.

Acknowledgments

The author would like to express sincere gratitude to Dr. Sarah Khan and Dr. Sam Barclay for their invaluable expertise, wholehearted guidance, and insightful comments, which greatly contributed to refining my thoughts. Additionally, my gratitude is extended to the participants whose willingness to engage in this long-term study. The research would not have been completed smoothly and successfully without their support and collaboration.

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*(s1), 1-26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. https://ci.nii.ac.jp/ncid/BB17915922

Caspi, T. (2010). *A dynamic perspective on second language development*. [Doctoral dissertation, University of Groningen]. s.n.

Costa, A. S., Dogan, I., Schulz, J. B., & Reetz, K. (2019). Going beyond the mean: Intraindividual variability of cognitive performance in prodromal and early neurodegenerative disorders. *The Clinical Neuropsychologist, 33*(2), 369-389. https://doi.org/10.1080/13854046.2018.1460735

De Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal, 92*(2), 166-178. https://doi.org/10.1111/j.1540-4781.2008.00712.x

De Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In *A dynamic approach to second language development: Methods and techniques* (pp. 5-24). John Benjamins Publishing Company.

De Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. Psychology Press.

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and cognition, 10*(1), 7-21. https://doi.org/10.1017/S1366728906002732

De Bot, K., Lowie, W., Thorne, S. L., & Verspoor, M. (2013). Chapter 10. Dynamic systems theory as a comprehensive theory of second language development. In *Contemporary approaches to second language acquisition* (pp. 199–220). John Benjamins Publishing Company. https://doi.org/10.1075/aals.9.13ch10

Dong, J. (2016). A dynamic systems theory approach to development of listening strategy use and listening performance. *System, 63*, 149-165. https://doi.org/10.1016/j.system.2016.10.004

Dörnyei, Z. (2014). Researching complex dynamic systems: 'Retrodictive qualitative modelling' in the language classroom. *Language Teaching*, *47*(1), 80–91. https://doi.org/10.1017/s0261444811000516

Dörnyei, Z., & Murphey, T. (2003). *Group dynamics in the language classroom*. Cambridge University Press.

Dörnyei, Z., MacIntyre, P. D., & Henry, A. (2014). Motivational dynamics in language learning. In *Multilingual Matters eBooks*. https://doi.org/10.21832/9781783092574

Drummond, A. (2018). Investigating the relationship between IELTS scores and receptive vocabulary size. *Journal of the Foundation Year Network, 1*, 113-125. https://jfyn.co.uk/index.php/ukfyn/article/download/13/23

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Eskildsen, S. W. (2012). The development of construction learning in a second language: The relation between chronological age and proficiency. *Applied Linguistics*, *33*(2), 131-149.

Fischer, K., Yan, Z., & Stewart, J. (2005). Adult cognitive development: Dynamics in the developmental web. *Handbook of developmental psychology*, 491–516. https://doi.org/10.4135/9781848608306.n21

Godwin-Jones, R. (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology, 22*(2), 8–27. https://doi.org/10125/44643

Harshbarger, B. (2007). Chaos, complexity and language learning. *Language Research Bulletin, 22*, 17-31.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. National Council of Teachers of English. https://ci.nii.ac.jp/ncid/BA19525648

Huynh, T. (2021). A complex dynamic systems approach to foreign language learners' anxiety in the emergency online language classrooms. *Computer-Assisted Language Learning Electronicl Journal*, *22*(3), 200-229.

Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *Modern Language Journal*, *92*(s1), 1–124. https://doi.org/10.1111/j.1540-4781.2008.00821.x

Kramsch, C. (2012). Why is everyone so excited about complexity theory in applied linguistics? *Mélanges Crapel, 33*, 9-24.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165. https://doi.org/10.1093/applin/18.2.141

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619. https://doi.org/10.1093/applin/aml029

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language development. *Applied Linguistics, 30*(4), 579-589. https://doi.org/10.1093/applin/amp043

Larsen-Freeman, D., & Cameron, L. (2007). Complex systems and applied linguistics. *International Journal of Applied Linguistics, 17*(2), 226-240. https://doi.org/10.1111/j.1473-4192.2007.00148.x

Lewis, M. (2002). Cultures of teaching: Voices from Vietnam. *ELT Journal, 56*(2), 146-153. https://doi.org/10.1093/elt/56.2.146

Lewontin, R. (2000). *The triple helix: Gene, organism, and environment*. Harvard University Press.

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. Seliger & M. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp. 217–243). Newbury House.

Lightbown, P., & Spada, N. (2006). *How languages are learned: An introduction to the main theories of first and second language acquisition (now in a New, Updated Edition)*. Oxford University Press.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33*(2), 135-158. https://doi.org/10.1111/j.1467-1770.1983.tb00532.x

Morrison, K. (2006). Complexity theory and education. In *APERA Conference*. Hong Kong.

Murakami, A. (2016). Modeling systematicity and individuality in non-linear second language development: The case of English grammatical morphemes. *Language Learning, 66*(4), 834-871. https://doi.org/10.1111/lang.12166

Ngo, K. T. (2025). EFL learners' perceptions of translanguaging in English classes at an HCM English centre. *International Journal of Language Instruction*, *4*(1), 50–64. https://doi.org/10.54855/ijli.25414

Nguyen, T. M. (2025). IELTS washback as a high-stakes test on student learning: A hierarchical modelling study at a Vietnamese university. *International Journal of TESOL & Education, 5*(1), 1–24. https://doi.org/10.54855/ijte.25511

Nguyen, T. T., & Pham, T. H. (2019). Vietnamese EFL learners' lexical acquisition: Variability, progress, and regression. *Journal of Language and Education, 5*(3), 45–62.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Longman Publishing Group.

Perone, S., & Simmering, V. R. (2017). Applications of dynamic systems theory to cognition and development: New frontiers. *Advances in child development and behavior, 52*, 43-80. https://doi.org/10.1016/bs.acdb.2016.10.002

Plotkin, V. (2006). *The language system of English*. Universal-Publishers.

Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 35*(2), 184-207. https://doi.org/10.1093/applin/amt013

Robinson, B. F., & Mervis, C. B. (1998). Disentangling early language development: Modeling lexical and grammatical acquisition using an extension of case-study methodology. *Developmental Psychology, 34(2)*, 363–375. https://doi.org/10.1037/0012-1649.34.2.363

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Basic Books.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of finish. *Applied Linguistics, 31*(4), 532-553. https://doi.org/10.1093/applin/amq001

Thelen, E., & Smith, L. B. (2006). Dynamic systems theories. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., pp. 258–312). Wiley.

Van Geert, P., & Steenbeek, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review, 25*(4), 408–442. https://doi.org/10.1016/j.dr.2005.10.003

Van Geert, P., & Steenbeek, H. (2008). A complexity and dynamic systems approach to developmental assessment, modeling and research. In K. W. Fischer, A. Battro, & P. Léna (Eds.), *The educated brain; Essays in neuroeducation* (pp. 71-94). Cambridge University Press.

Van Gelder, T., & Port, R. F. (1998). It's about time: An overview of dynamical approach to cognition. *The MIT Press eBooks*. https://doi.org/10.7551/mitpress/4622.003.0002

Verspoor, M., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development: Methods and techniques*. John Benjamins Publishing Company.

Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, *92*(2), 214–231. https://doi.org/10.1111/j.1540-4781.2008.00715.x

Yang, W., & Sun, Y. (2015). Dynamic development of complexity, accuracy and fluency in multilingual learners' L1, L2, and L3 writing. *Theory and Practice in Language Studies, 5*(2), 298-308. https://dx.doi.org/10.17507/tpls.0502.09

Biodata

Dinh Thao Nguyen is a lecturer of the Faculty of Foreign Languages at Dong Nai University, and she has over eight-year experience in teaching English to non-English and English major students. She got her master's degree in English Language Teaching at Nottingham Trent University, England. Her research interests mostly lie in Dynamic System Theory (DST) and Vocabulary Research.

Teaching EFL Vocabulary through Analyzing the Structure of Words Coupled with Using Crossword Puzzles

Yu-Chi Yang[1*]

[1]Department of Foreign Languages, Fooyin University, Kaohsiung City, Taiwan
[*]Corresponding author's email: yuchiyang0813@gmail.com
[*] https://orcid.org/0009-0008-9633-2988
https://doi.org/10.54855/ijte.25523

Received: 05/11/2024 Revision: 04/03/2025 Accepted: 24/03/2025 Online: 23/04/2025

ABSTRACT

In the EFL milieu, plenty of studies have reported the usefulness of some vocabulary-building techniques. Among them, teaching incorporating either the crossword or mnemonics has been widely implemented in various language settings. Empirical evidence supports the value of such a teaching method. Nonetheless, using both the devices in tandem deserves attention and further investigation. In the current research, two groups of EFL majors in a junior college in Taiwan were recruited to receive either traditional lecture-based instruction or teaching aided by both the crossword and mnemonics. Data analysis demonstrates significant inter-group differences in the learning outcome, favorable to the latter group, and the questionnaire responses exhibit the students' approval of this experimental technique. The findings of the research mean more to EFL teaching in Taiwan, as Mandarin Chinese, the medium of instruction in schools, is by far different from English, making it more difficult to learn English as L2.

Keywords: mnemonics, word structure diagram, crossword puzzle, vocabulary acquisition, memory

Introduction

English is a "global lingua franca" used in international politics, entertainment, air traffic control, academia, trade, diplomacy, and social media (Mauranen, 2010, p. 6). In Taiwan, English is the most widely learned foreign language. Public policy wise, the government recognizes the crucial role of English in international and intercultural communication, and English-language teaching is implemented at all levels of education (Ministry of Education, 2018; Rüdiger et al., 2023). In practice, English is a high-stakes subject of study for most students. The outcome of their performance on this subject has a significant impact on their chance of being admitted to a certain major in a certain university. English skills are also a qualification for some appointments in the government and private sectors (104 Corporation,

CITATION | Yang, Y. C. (2025). Teaching EFL Vocabulary through Analyzing the Structure of Words Coupled with Using Crossword Puzzles. *International Journal of TESOL & Education*, *5*(2), 57-74. DOI: https://doi.org/10.54855/ijte.25523

2021; Yam News, 2023).

That being said, opportunities for Taiwanese learners to use the language are limited. There are reasons for and consequences of this awkward situation. Politically, Taiwan is quite isolated in the international community, limiting mass popular access to English as a medium of international communication and thus rendering this language less appreciated (Republic of China (Taiwan), 2018). On a smaller scope, classroom instruction in Taiwan, which is influenced by social expectations and a rigid educational system, prioritizes test-taking skills over proficiency. Young learners' mentality toward learning English is often short-sighted, i.e. striving to fulfill an imminent purpose such as passing an exam through rote memorization (Chen et al., 2020; Huang, 2014; Liao, 2004; Yeh & Wang, 2004), which could explain the omnipresence of private learning centers (Rüdiger et al., 2023). This focus often leads to insufficient effort being directed toward improving learners' ability to comprehend and use the language. For Taiwanese English teachers, as expected, learners' interest and motivation are frequently secondary concerns. Additionally, the popularity of e-learning (Tran & Nguyen, 2022) has enticed many learners to believe that watching videos on social media can allow them to improve their vocabulary and English proficiency. This unstructured approach to learning exacerbates these learners' inability to successfully improve their proficiency levels. This observation of learners' preference for technology as a learning aid is in agreement with the findings of Le and Trinh (2024). And yet, the more concrete and obvious consequence of these scenarios is learners' subpar performance on international English language proficiency exams, such as TOEFL and TOEIC. Recently, Taiwanese TOEFL test takers performed below the average on all four components: Reading, Listening, Speaking and Writing (ETS, 2023). On the Test of English for International Communication (TOEIC®) program, Taiwanese test takers' level of proficiency reveals their inadequate vocabulary competence coupled with marginal written expression and simple conversation skills (Chun, 2024).

The researcher of this study, in his interactions with his students, has noted that young learners often reject thoughtfully designed approaches, even when the less thoughtful methods they adopt fail to produce desired outcomes. Students also often complain about vocabulary being difficult to learn and easy to forget, reflecting their frustration with the learning process.

On the basis of his professional experience, the researcher of this study designed an intervention aimed at promoting learning engagement and, by elucidating alternative learning strategies, enabling learners to access other learning resources. This intervention integrated the Fish-skeleton Vocabulary Learning Diagram (FSVLD) (Appendix 1) and crossword puzzles (Appendix 2). The FSVLD utilizes was used to illustrate word composition and assist with identification of the meanings of word components. After completing lessons focused on the FSVLD, the students applied their knowledge to solve crossword puzzles.

The rationale for this study can be elucidated from three perspectives. First, in EFL settings, the use of the FSVLD is a new practice. Yang and Wang (2006) used this device as an aid for writing essays as well as for quickly grasping the ideas of a writing. Afterwards, Yang and Wang (2014) used it to help students acquire vocabulary knowledge. However, the effectiveness of this device in achieving the goal of vocabulary acquisition and memory has not been validated.

Second, comparative linguistics justifies revisiting crossword puzzles. In contrast to the FSVLD, the crossword puzzles have been widely used in EFL classrooms and their effectiveness in facilitating vocabulary acquisition has been confirmed in previous studies (Alda & Wati, 2021; Keshta & Al-Faleet, 2013; Lestari & Yulia, 2018; Njoroge et al., 2013; Merkel, 2016; Mustika et al., 2022; Orawiwatnakul, 2013). However, the wide use of crossword puzzles is not a reason at all why the researcher of this study jumped on the bandwagon of investigating the potential of this device. Rather, it is comparative linguistics that prompts him to do this study. The process of acquiring L2 vocabulary is complex (Ryan, 1997; Schmitt & McCarthy, 1997). It should be noted that how similar L1 and L2 are decides how easy or difficult it is to pick L2 vocabulary or to learn L2 more generally. More similarity makes this process easier (Schmitt & McCarthy, 1997), and Richards (1976) implies that the level of such similarity is decided by several features of words of a language. Mandarin Chinese is the first language used in Taiwan, and it is the predominant medium of instruction in all levels of education (Rüdiger et al., 2023). There exist immense differences between Mandarin Chinese and English, and the gap must be larger than those between English and other alphabetical languages. These differences permeate various domains of language: orthography, phonology, syntax, and semantics. Mandarin Chinese doesn't belong to the alphabetical system. Rather, a Chinese character is made up of several strokes, for example, "快樂〞, pronounced "kwai lur" and literally meaning "happiness." Each of these two characters is formed by a number of strokes. As to syntax, in Chinese, there is no subject-verb inversion. The verb doesn't have different forms, so it is virtually impractical to compose verbal phrases, and EFL learners often have difficulty using verb tense correctly. Collocations also cause trouble for Taiwanese learners. Take the word 'wait' as an example. A learner may know its meaning. But the phrases 'wait on' and 'wait for' may be confusing or misleading. The effect of the inherent gap existing between Mandarin Chinese and English in relation to their linguistic features could be augmented by cultural differences (Merkel, 2016).

Third, previous studies on the usefulness of the crossword puzzle focused on its impact on vocabulary acquisition; empirical evidence of its impact on memory is scarce and inconsistent. Furthermore, Taiwanese teachers and scholars have not done research on the effectiveness of the crossword puzzle, either in vocabulary acquisition or in vocabulary memory. Therefore, to validate the possible usefulness of the FSVLD, obtain empirical evidence of the applicability of the crossword puzzle in another lingual and cultural context, and validate the effect of the experimental intervention on memory, justify this study.

To test the effectiveness of this intervention, the researcher of this study conducted a quantitative study. One cohort of students received this intervention, and another received traditional lecture-based instruction. The performance of these cohorts was statistically analyzed to address two research questions:

1. Can FSVLD-centered explanations coupled with the use of crossword puzzles help EFL learners improve their ability to acquire vocabulary?
2. Can FSVLD-centered explanations coupled with the use of crossword puzzles help EFL learners remember acquired vocabulary?

Literature Review

A considerable number of scholars have highlighted the crucial role of vocabulary in expressing an individual's meaning. For example, linguist Peter Funk asserted that words are central to cognitive processes (Reader's Digest, 1983). Similarly, a sociolinguist named Wilkins emphasized vocabulary's pivotal role in language, noting, "While without grammar, very little can be conveyed, without vocabulary, nothing can be conveyed" (1985, p. 111). Wilkins (1985) also suggested that vocabulary knowledge profoundly influences an individual. In the same vein, King (2010, p. 5) described vocabulary as a core component of expression, relating it to the "bricks" of a valid piece of writing, and Richards and Renandya (2002, p. 255) referred to vocabulary as "a core component of language proficiency," emphasizing that it provides "much of the basis for how well learners speak, listen, read, and write".

The role of vocabulary in the learning process has been widely studied. Numerous studies have linked vocabulary knowledge to academic success, in addition to the development of other language skills. For example, Nation (1990) asserted that the necessity of vocabulary knowledge for effective learning cannot be overestimated. The logic of this causation is supported by some sources of information. Orawiwatnakul observed that vocabulary is "a key basis on which reading achievement depends" and is crucial to meaning and comprehensible expression (2013, p. 414). According to Karami and Bowles (2019) and Nam (2010), vocabulary can impact an individual's cultural understanding, besides writing, reading, listening, and speaking. Additionally, Keshta and Al-Faleet (2013) discovered that vocabulary determines reading comprehension levels and thereby influences academic performance. The National Institute of Child Health and Human Development [NICHD] summarized the close relationship between vocabulary and academic development as follows: "Vocabulary is important for reading to learn as well as learning to read (NICHD, 2020, p. 22)."

As individuals of diverse backgrounds began learning English (King, 2010) and EFL was recognized as an academic discipline (Faraj, 2015) in the second half of the 20th century, researchers and EFL teachers began developing strategies to enhance vocabulary acquisition and memory (Wei, 2007). One such strategy involves mnemonics (Agnes & Srinivasan, 2024a, 2024b; Amiryousefi & Ketabi, 2011; Farjami, 2007; Hulstijn, 1997; Kurniarahman, 2023; Pillai, 2017). Farjami (2007) claimed that mnemonic devices have value in teaching vocabulary and helping with memory of vocabulary knowledge. Agnes and Srinivasan (2024a, 2024b) and Amiryousefi and Ketabi (2011) emphasized the ability of mnemonic to link new information to existing knowledge and thus help a learner retrieve cues stored in the brains and cited visual imagery as one of the essential mechanisms through which mnemonic devices function. Pillai (2017) observed that mnemonics, which can be presented in visual, physical, or other forms, provide substantive stimulation, create vivid impressions, attract attention, and facilitate memory. Kurniarahman (2023) underscored the positive effect of mnemonic devices on vocabulary memory, possibly owing to their potential to engage students in learning.

Inspired by previous literature on mnemonics (Amiryousefi & Ketabi, 2011; Farjami, 2007; Hulstijn, 1997), Yang and Wang (2014) developed the FSVLD as a teaching aid. This diagram

is shaped like a fish skeleton and comprises three parts: head, trunk, and tail, representing the prefix, root, and suffix of a word, respectively (Yang & Wang, 2014). According to Yang and Wang (2014), when thoughtfully implemented, this diagram can aid learners in developing idiomatic usage, knowledge of antonyms and synonyms, spelling, and the habit of associating concepts. As this diagram demonstrates three parts of a word, it enables the students to quickly form an image of the meaning contained in the word and how it functions that way. As elucidated in literature on mnemonics, in comparison with verbal description, visual imagery is more capable of drawing attention and creating an engaging learning process. From a cognitive perspective, visual information is more easily processed and stored by the human brain, and more capable of creating mental representations and bringing about meaningful connections (Alabi, 2024). In addition, gaining an understanding of the composition of words is an effective method for learning vocabulary (Ellis, 1997; Li, 2009). This benefit can be reasoned as a result of reduced workload. Instead of learning a myriad of words separately, learners only have to remember a lot fewer units of information. With such knowledge, it would be possible for them to guess the meaning of new words or to memorize the learned words more effectively.

Using the crossword puzzle is another strategy that has been validated as effective in enhancing vocabulary acquisition (Alda & Wati, 2021; Keshta & Al-Faleet, 2013; Lestari & Yulia, 2018; Merkel, 2016; Mustika et al., 2022; Njoroge et al., 2013; Orawiwatnakul, 2013) for several reasons. First, solving crossword puzzles is fun and enjoyable. Keshta and Al-Faleet (2013) observed that learners enjoy learning methods that are engaging and that enable "interactive, exciting, and fun learning," highlighting crossword puzzles' potential for diversifying classroom activities and establishing a relaxing learning atmosphere (p. 47). Merkel (2016) noted that solving vocabulary crosswords is fun and renders learning satisfying and stimulating. It is worth noting that when these educators had a favorable experience with crossword-aided teaching, it had already been used in such disciplines as communication, health, psychology, reading, and sociology classes because it is conducive to a more cheering learning experience (Childers, 1996).

Second, crossword puzzles inspire critical thinking. By Ausubel (1963), Mayer (2002), and Yunianta et al. (2012), active thinking is instrumental in achieving meaningful learning. Ayto (1990, p. 422) observed that "arduous reasoning" is essential for solving puzzles. Similarly, Childers (1996) recognized the crossword puzzle's ability to evoke critical thinking and imagination. Additionally, Krashen (1985) suggested that unchallenging learning, that is, learning that is excessively easy or boring, often fails to inspire learners, whereas "i + 1" (input slightly higher than the learner's comprehension level) promotes effective learning.

Third, crossword-assisted teaching encourages engagement and the manipulation of concepts and ideas. Schmitt and McCarthy (1997) described solving crosswords as a cognitive process involving recalling and manipulating words that leads to deep engagement. Moreover, Gairns and Redman (1999) noted that solving a crossword puzzle requires application of a range of information, including grammar and meaning. Finally, studies have demonstrated (Burston, 2005; Merkel, 2016; Schmitt & McCarthy, 1997) that language learners prefer searching for meaning in word clusters to enhance the acquisition of new vocabulary. As solving crossword

puzzles requires the player to read clues, this activity exposes learners to substantive and correct word strings, facilitating thorough learning.

On the strength of the potential impact of the FSVLD and the validated benefits of crossword puzzles, it is not too far-fetched to say that the intervention implemented in this current study can facilitate meaningful learning, an educational concept proposed by Ausubel (1963). Meaningful learning involves active thinking, motivation, problem-solving, practice, and a sense of achievement (Yunianta et al., 2012). Mayer (2002) suggested that meaningful learning has lasting effects and helps learners solve problems beyond those presented in learning tasks.

Learners and teachers are jointly responsible for achieving meaningful learning. Teachers' adjustments to curriculum pedagogy also impact on students' meaningful learning experiences (Vu et al., 2020). Huang (2005), Karami and Bowles (2019), Nemati (2009), and Pillai (2017) have indicated that learners must engage and persist in their efforts, and teachers must create a motivational atmosphere and employ diverse instructional methods to support learners. These views are consistent with that of Wilkins (1985), who observed that recognizing the practical value of learning a second language — using it to influence the environment — increases learners' intent to study.

Commitment, motivation, and guidance are crucial to knowledge acquisition and retention. Therefore, teachers must create and maintain a learning environment that sustains learners' momentum and guides them appropriately. The pedagogical concepts buttressing the effectiveness of the intervention in the present study are the Zone of Proximal Development (ZPD) (Vygotsky, 1978) and instructional scaffolding (Lipscomb et al., 2010; Wood et al., 1976). By these concepts, instruction should progress from easy tasks to more difficult tasks, with teachers providing support on the basis of students' competence levels until the students achieve curriculum objectives.

Methods

This study was conducted over 23 weeks of the 2022-23 academic year. The duration spanned two consecutive semesters (including a four-week winter break in-between). During this period, students taking an English vocabulary enhancement course were divided into experimental and control groups. The experimental group received the experimental intervention, and the control group received traditional lecture-based instruction. The Methods contain five major sections: research design, participants, study instruments, intervention, and data processing. They are enumerated as follows.

Research Design

This study adopted a quasi-experimental design model, shown in Table 1. Initially, a pretest was administered to both groups of students. After the pretest, the larger cohort, designated as the experimental group, received the intervention and completed a questionnaire. Following the intervention, both the experimental and the control groups completed a posttest and a follow-up test. These tests comprised the same set of questions as those on the pretest.

Table 1

Quasi-experimental design model

Group	Pretest	Experimental treatment[1]	Posttest	Follow-up test	Response to questionnaire
Experimental	Yes	Yes	Yes	Yes	Yes
Control	Yes	No	Yes	Yes	No
Experimental	Yes	Yes	Yes	Yes	Yes

Note. [1]Fish-Skeleton Vocabulary Learning Diagram + Crossword Task Activity.

Participants

This study involved 71 third-year students from a 5-year junior college EFL program (equivalent to the senior year in high school) enrolled in the course "English: Vocabulary and Etymology." The participants were divided into two groups: 36 in the experimental group and 35 in the control group. Statistical analyses were conducted using SPSS version 28. A pretest with 60 questions revealed nearly identical vocabulary proficiency between the two groups, with mean scores of 26.22 and 26.37, respectively. Statistical tests (t-test and Levene's test) confirmed no significant difference between the groups ($p = 0.73$). The experimental group was selected due to its larger size for reasons of convenience and practicality. Additionally, the study involved the researcher, three faculty members, and 55 EFL students who were not part of the main groups.

Study Instruments

The study utilized various tools, including a textbook, a learning achievement test, the FSVLD, crossword puzzle worksheets, and a learning attitudes questionnaire. The textbook, Reading Explorer (3rd edition, 2019), published by Cengage Learning, offered a wide range of reading passages aimed at broadening and deepening students' vocabulary. A 60-question learning achievement test embedded in the textbook was administered to both cohorts as a pretest, a posttest, and a follow-up test conducted four weeks after the instruction.

The purpose of FSVLD is to explain word composition, clarify the meanings of word elements, and encourage students to create new words using prefixes, roots, or suffixes, thereby enhancing their understanding of vocabulary. These activities prepared students for subsequent crossword puzzle tasks. The reason for using a fish skeleton diagram for vocabulary teaching is that, in etymology, words are divided into three parts: prefix, root, and suffix, which correspond to the head, trunk, and tail of a fish. Additionally, using a visual representation helps students create a mental image when learning words, enabling them to retain vocabulary for a longer time.

To develop the crossword puzzle worksheets, the researcher collaborated with three faculty members, using a free crossword puzzle generator from The Teacher's Corner website. Target words and their definitions, selected from the textbook, were input into the software to create 12 worksheets, each containing 18 or 19 target words. Of these, one was used for demonstration, nine as quizzes, and two for practice. The accuracy of these worksheets was verified by the faculty members.

A learning attitudes questionnaire (Appendix 3), designed by the researcher, was distributed to the experimental group after the instruction cycle to evaluate the effectiveness of the FSVLD and crossword puzzles in enhancing learning and retention. The questionnaire, comprising 29 items, was scored on a 5-point Likert scale. Exploratory factor analysis using principal component analysis identified three factors—memory, acquisition, and sense of achievement—which explained 64.434% of the variance. The Kaiser–Meyer–Olkin value was 0.807 ($p < .001$), confirming the data's suitability for factor analysis. The questionnaire exhibited excellent reliability, with Cronbach's α at 0.963.

Intervention

After dividing the students into groups, the researcher of this study implemented two distinct teaching approaches. The control group followed a lecture-based method, focusing on word explanations, grammar reviews, and discussions on word origins or related stories. In contrast, the experimental group combined FSVLD instruction with a 20-minute crossword activity during each 50-minute class. Other activities mirrored those of the control group but were delivered in a more concise and faster manner.

To begin, the researcher of this study demonstrated solving a crossword puzzle to familiarize the experimental group with the process and encourage participation. During the intervention, the group completed nine crossword quizzes and collaborated with the researcher of this study in two class sessions to solve puzzles together. As Zitouni et al. (2021) emphasized, repetitive and engaging exercises over time significantly enhance vocabulary retention. After the posttest, the experimental group completed a feedback questionnaire.

Data Processing

The researcher of this study utilized various statistical methods to analyze data from the experimental group's crossword quizzes, as well as both groups' posttest and follow-up test results. First, the experimental group's crossword quiz scores were recorded and standardized as z-scores. Second, t-tests compared pretest and posttest scores of both groups to identify significant differences. Third, follow-up test results were examined to evaluate vocabulary retention. Fourth, Pearson correlation analysis assessed the experimental treatment's impact on posttest outcomes. Lastly, questionnaire responses were analyzed to provide additional insights, complementing the study's findings.

Findings

The data analysis yielded the following insights into the first research question and into the second research questions, and the statistical treatment of the responses to the questionnaire adds credibility to these insights.

Four data sets were analyzed to answer the first research question, with the statistical analysis shown in Table 2, Table 3, and Table 4. The data sets comprise (1) The Experimental Group's Performance on the Crossword Quizzes, (2) Both Groups' Performance on the Pretest and the Posttest, (3) The Groups' Performance on the Posttest, and (4) The Correlation Between the Experimental Group's Performance on the Crossword Quizzes and on the Posttest.

Data Sets Pertinent to the First Research Question

The Experimental Group's Performance on the Crossword Quizzes

The scores of the experimental group on the nine quizzes demonstrate a consistent upward trend throughout the instruction cycle. The group achieved an overall z-score of 55.26, with scores of 38.12 on the first quiz, 53.36 on the fifth quiz, and 77.78 on the ninth quiz. The extent of improvement from the first to the fifth quiz was 39.9%, that from the fifth to the ninth quiz was 45.8%, and that from the first to the ninth quiz was 104%, indicating substantial and sustainable progress in completing the crossword tasks.

Both Groups' Performance on the Pretest and the Posttest

The experimental group had an average of 26.22 correct answers on the pretest and 35.58 on the posttest, and the control group had 26.37 correct answers on the pretest and 24.94 on the posttest. Paired sample *t*-tests were conducted to examine variations in each group's mean scores. The results for the experimental group yielded a t value of −9.194, with $p < .001$, shown in Table 2, indicating a significant difference in their performance between the two tests. By contrast, the results for the control group yielded a t value of 1.274, with a p value of 0.211, shown in Table 3, indicating no significant differences in their performance between the two tests.

The Groups' Performance on the Posttest

The groups' performance on the posttest was also compared. The data were subjected to independent sample *t*-tests, yielding $t = -6.378$ and $p < .001$, shown in Table 4, indicating significant differences in the performance of the experimental and control groups.

The Correlation Between the Experimental Group's Performance on the Crossword Quizzes and on the Posttest

The experimental group scored an average of 10.51 (out of 18 or 19 questions on each quiz) on the nine quizzes. Pearson correlation analyses were subsequently conducted to assess whether the group's performance on the quizzes was correlated with their performance on the posttest. The result, $r = 0.64$ and $p < .001$, indicates a moderate correlation between these data points.

The analysis of the above four data sets highlights that the intervention in this study led to significantly different learning outcomes between the experimental and control groups. Throughout the instructional period, the experimental group consistently outperformed the control group in vocabulary acquisition, demonstrating the intervention's effectiveness and providing a clear benefit to those who received the specialized instruction.

Table 2

Paired sample t-test of the experimental group's performance on the pretest and the posttest

Measuring	Mean	SD	T	df	2-tailed p
Pre-test	26.22	6.109	−9.194	35	<.001
Post-test	35.58				

Note. ***$p < .001$.

Table 3

Paired sample t-test of the control group's performance on the pretest and the posttest

Measuring	Mean	SD	T	Df	2-tailed p
Pre-test Post-test	26.37 24.94	6.635	1.274	34	0.211

Table 4

Independent sample t-test results for the posttests of the control and experimental groups

Group	N	Mean	SD	T	2-tailed p
Control Experimental	35 36	24.94 35.58	6.637 7.389	−6.378	<.001

Note. ***$p < .001$.

To answer the second research question, two data sets were analyzed, with the statistical analysis shown in Table 5, Table 6, and Table 7. The data sets comprise (1) Both Groups' Performance on the Posttest and the Follow-up Test. and (2) The Groups' Performance on the Follow-up vs Their Performance on the Previous Tests.

Data Sets Pertinent to the Second Research Question

Both Groups' Performance on the Posttest and the Follow-up Test

Specifically, the groups were compared in terms of the difference in performance on the posttest and the follow-up test. The experimental group scored an average of 35.58 on the posttest and 34.97 on the follow-up test. These data were subjected to a *t*-test, which yielded $t = 0.828$, with $p = .413 > .05$, indicating no significant differences between the experimental group's performance on the two tests. By contrast, the result for the control group on the posttest was 25.04[2] and that on the follow-up test was 19.87. These data were also subjected to a *t*-test, yielding results of $t = 6.443$ and $p < .001$, indicating significant differences between the control group's performance on these two tests.

The Groups' Performance on the Follow-up vs Their Performance on the Previous Tests

Table 5

Mauchly's spherical test

Within-subject effects	' W	Chi-square test	df	Sig	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	.869	4.776	2	.092	.884	.928	.500

The researcher of this study also used a repeated-measure analysis of variance to compare the groups in terms of their performance on the follow-up test versus their performance on the pretest and the posttest. The results revealed that the experimental group earned a mean score of 34.97 (standard deviation [SD] = 6.92) on the follow-up test and 26.22 (SD = 6.30) and 35.58 (SD = 7.39) on the pretest and posttest, respectively. Paired analysis, conducted using generalized linear models and repeated measurements, yielded a Mauchly's W coefficient of .869 ($X^2 = 4.776$, $p = .092$), passing Mauchly's spherical test, shown in Table 5.

Table 6

Test of intraparticipant effects

Source		Type III sum of square	df	mean square	F	Sig.
Time	Finding Sphericity Assumed	1974.796	2	987.398	69.872	<.001
	Greenhouse-Geisser	1974.796	1.768	1116.800	69.872	<.001
	Huynh-Feldt	1974.796	1.855	1064.363	69.872	<.001
	Low-bound	1974.796	1.000	1974.796	69.872	<.001
Error (Time)	Error Sphericity Assumed	989.204	70	14.131		
	Greenhouse-Geisser	989.204	61.889	15.983		
	Huynh-Feldt	989.204	64.938	15.233		
	Low-bound	989.204	35.000	28.263		

A test of intraparticipant effects attained F = 69.872 with $p < .001$, reaching a level of significance, shown in Table 6.

Table 7

Paired comparison results

(I) Time	(J) Time	MD (I-J)	SE	Significance	95% confidence interval for the difference	
					LL	UL
1	2	−9.361*	1.018	<.001	−11.428	−7.294
	3	−8.750*	.880	<.001	−10.536	−6.964
2	1	9.361*	1.018	<.001	7.294	11.428
	3	.611	.738	.413	−.887	2.110
3	1	8.750*	.880	<.001	6.964	10.536
	2	−.611	.738	.413	−2.110	.887

Note. *. The difference in means is significant at the .05 level.

The experimental group scored higher on the posttest than on the pretest, with a significant difference of 9.361 and a standard error (SE) of 1.018 ($p < .001$). This group also performed better on the follow-up test than on the pretest, with a significant difference of 8.750 and SE of 0.88 ($p < .001$). The group's performance on the posttest was superior to that on the follow-up test, with a difference of .611 and SE of .738, but this difference was not significant, with $p = .413$, shown in Table 7. These results indicate that the experimental intervention enhanced vocabulary memory. However, this pattern of changes didn't happen to the control group. As shown in Table 3, the difference in control group's performance on the pretest and the posttest is not significant, but this group performed worse on the follow-up than on the posttest, with the difference reaching a level of significance.

The analysis of the above two data sets demonstrates that the intervention of this current study produced significantly different learning outcomes between the groups, progressing from the inception of the instruction to the administration of the follow-up test, favorable to the experimental group.

Finally, on learning attitudes, the scores given to the questionnaire items were averaged,

attaining 4.2/5.0, and a Pearson Correlation analysis was conducted to examine the correlation between learning attitudes and performance on the follow-up, yielding a Pearson's r of 0.73, with $p < .001$. These statistics speak of the students' high opinion of the experimental intervention and show that the intervention is effective in inspiring active search for clues for executing a task, sustaining engagement, and bringing about a sense of satisfaction. The evidence can be drawn from the responses to some of the questionnaire items: "When completing a crossword task, I approach spelling a complete target word like solving a problem, which gives me a sense of satisfaction.," "Solving crossword puzzles using my knowledge of the affixes and roots of words enables me to form a vivid impression of the words I have learned and remember them later." and "My improved vocabulary, confidence in using English, and grades after taking this course have enhanced my sense of accomplishment in learning English." Furthermore, a Pearson's Coefficient was conducted to examine the relationship of learning interest to success. A Pearson's $r = 0.245$, with $p = .150$, indicates a weak correlation between these two criteria.

Discussion

The effectiveness of the experimental intervention warrants consideration. The results of the current study evidence the potency of this teaching strategy to enhance vocabulary acquisition and memory. Overall speaking, teaching through using the FSVLD and crossword puzzles in sequence effectively holds students' attention, fosters engagement, and provides opportunities for practice, rendering the learning process more enjoyable and more capable of enhancing vocabulary acquisition and memory. These benefits are primarily attributable to the alignment of this teaching strategy with the principles of meaningful learning (Ausubel, 1963; Mayer, 2002; Yunianta et al., 2012), which suggests that active thinking, meaning comprehension, motivation, and practical application facilitate consolidation of learning outcomes and improve an individual's ability to solve real-life problems.

Speaking separately, the impact of the FSVLD can be explained from three perspectives. First, it has the potency to allow students to explore word structures, thus enhancing their understanding of the words. After gaining knowledge of prefixes, roots, and affixes, the students expanded their vocabulary by modifying word components. Second, since a word is divided into three parts, each with an endowed meaning, instead of comprising a larger number of separate alphabetical letters, the students were able to understand the word better and can maintain the momentum for learning owing to the reduced workload. More importantly, as the vocabulary learning diagram visually presents the composition of words, it is more capable of attracting attention and reducing boredom. The diagram presents visual stimulation and facilitates forming concrete imagery of words, thus conducive to enhanced memory. The finding along with the explanations for it is in agreement with the linguistic research of Ellis (1997) and research on mnemonics (Agnes & Srinivasan, 2024a, 2024b; Amiryousefi & Ketabi, 2011; Hulstijn, 1997; Kurniarahman, 2023; Pillai, 2017).

On the positive impact of the crossword puzzle on vocabulary acquisition, it can be attributed to the opportunity to practice and to a sense of purpose as a result of solving a problem. In a

crossword game, the students apply their knowledge and experience how the learned words function in real scenarios. Doing or manipulating not only solidifies learning but leads to satisfaction. In this regard, the outcome of this study agrees with those of a number of previous studies (Alda & Wati, 2021; Keshta & Al-Faleet, 2013; Lestari & Yulia, 2018; Merkel, 2016; Mustika et al., 2022; Njoroge et al., 2013; Orawiwatnakul, 2013). However, it doesn't align with the study outcome of Puspita and Sabiqoh (2017), which failed to prove the positive relationship of crossword puzzles to vocabulary learning.

Why the intervention implemented in this study has made a significantly positive impact on vocabulary memory entails complexity. In some previous research (Keshta & Al-Faleet, 2013). Puspita and Sabiqoh (2017), no significant impact of the crossword puzzle on vocabulary has been confirmed. Therefore, the result of this study may be attributed to the potency of the FSVLD. Or it may be linked to the effect of using two devices in sequence. This teaching activity is captivating, enlightening, pragmatic, and engaging, because the vocabulary learning diagram is capable of effecting visual stimulation and attracting attention and the crossword puzzle inspires thinking and doing. The effectiveness of this sequence of activities is supported by research regarding how the neural system consolidates memory. As pointed out by Hong (2022), conscious learning that has been put into practice has a greater chance of altering the connection of neural circuits, engendering long-term memory.

Conclusions and Suggestions

The findings of this current study contain four major points: (1) The teaching intervention has a significant positive impact on vocabulary acquisition, (2) The teaching intervention has buttressed learning endeavors, motivating learners to remain on track to attain the goal, (3) The teaching intervention has a significant positive impact on memory of vocabulary knowledge, and (4) The students accept the teaching intervention and have benefited from it. Positive answers have been given to both the research questions.

The current study is meaningful in a number of aspects. First, it sheds light on the credibility and validity of the usefulness of the crossword puzzle in EFL vocabulary learning. Second, it provides empirical evidence of the value of the crossword puzzle in achieving memory of vocabulary knowledge if this instructional tool is used in a proper context, such as coupled with another device. Third, it offers an insight into the potential of a vocabulary learning diagram. Finally, and most importantly, the findings of the study are not consistent with previous research on the effect of interest in learning. For example, Isangedighi (1997) reported a strong association between young learners' academic performance and their interest in learning and study habits. However, the current study shows, although the students started with a low initial interest, they adhered to their enthusiasm throughout the learning cycle and performed better than their counterparts in the Control group in the end. This suggests that students' success may be attributable to the instructional intervention.

There is no versatile, or the best, teaching strategy that is universally useful. Rather, numerous conditions must be taken into consideration when designing and implementing teaching strategies (Alabi, 2024; Dang & Tong, 2024; Kurniarahman, 2023; Le & Trinh, 2024; Pillai,

2017). Nonetheless, this study, along with the findings, offers an incentive to think up and implement alternative teaching techniques, especially for a literal and cultural context like Taiwan, where the first language is way different from English and still the educators and young learners are stuck in a mindset, striving to find a shortcut to attaining the imminent goal.

Since the current study investigated the combined effect of two instructional devices, there is a need to explore them separately. Further explorations for additional evidence of the effectiveness of the FSVLD are strongly suggested. Also deserving are investigations of the impact of the crossword puzzle on memory of vocabulary knowledge and of the influence of learning attitudes on outcomes. Whichever action is taken, every effort should be made to get access to a large sample of study. When conditions permit, a sample composed of learners from various disciplines (Le & Trinh, 2024). or involving instructors (Dang & Tong, 2024), should be able to safeguard a reliable outcome and increase the generalizability of the outcome.

Notes

1. Fish-skeleton Vocabulary Learning Diagram + Crossword Task Activity.
2. The size of the Control group was reduced to 23 when the second semester started owing to attrition. This sample size was applied when comparing the group's performances on the posttest and the follow-up test.

References

104 Corporation. (2021). *Global HRM & TOEIC*. 104 Corporation.

Agnes, D., & Srinivasan, R. (2024a). Redefining the association between memory, mnemonics and vocabulary acquisition—Reviewing paradigms in research. *Journal of Language Teaching and Research, 15*(3), 697–706. https://doi.org/10.17507/jltr.1503.02

Agnes, D., & Srinivasan, R. (2024b). Fostering vocabulary memorization: Exploring the impact of AI-generated mnemonic keywords on vocabulary learning through Anki flashcards. *World Journal of English Language*, *14*(2), 433–451.

Alda, J. N., & Wati, S. (2021). Enhancing learners' vocabulary acquisition by crossword puzzle game. *JL3T (Journal of Linguistics, Literature and Language Teaching), 7*(1), 25–40. https://doi.org/10.32505/jl3t.v7i1.2820 2

Amiryousefi, M., & Ketabi, S. (2011). Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching and Research, 2*(1), 178–182. https://doi.org/10.4304/jltr.2.1.178-182

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. Grune & Stratton.

Ayto, J. (1990). *Dictionary of word origins*. Arcade Publishing.

Burston, J. (2005). *Theoretical foundations of crossword puzzle usage in foreign language*

vocabulary acquisition. In Conference paper published at UNTELE, Academia.

Chen, F., Kao, S. M., & Tsou, W. (2020). Toward ELF-informed bilingual education in Taiwan: Addressing incongruity between policy and practice. *English Teaching & Learning*, *44*(2), 175–191. https://doi.org/10.1007/s42321-020-00055-1

Childers, C. D. (1996). Using crossword puzzles as an aid to study sociological concepts. *Teaching Sociology*, *24*(2), 231–235. https://doi.org/10.2307/1318816

Chun, S. (2024). *TOEIC® listening & reading test.* Chun Shin Limited.

Dang, T. M. T., & Tong, T. M. H. (2024). Instruction-giving in EFL classrooms: A case of a university in Viet Nam. *International Journal of TESOL & Education, 4*(2), 46–61. https://doi.org/10.54855/ijte.24423

Davis, T. M., Shepherd, B., & Zwiefelhofer, T. (2009). Reviewing for exams: Do crossword puzzles help in the success of student learning? *The Journal of Effective Teaching*, *9*(3), 4–10.

Derer, Ö. K., & Berkant, H. G. (2019). The effect of puzzle-based learning on secondary school students' attitudes and their self-efficacy beliefs in English lesson. *Journal of Theoretical Educational Science*, *13*(1), 205–231. https://doi.org/10.30831/akukeg.564792

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–139). Cambridge University Press.

ETS. (2023). *TOEFL iBT® test and score data summary 2022.* https://www.ets.org/pdfs/toefl/toefl-ibt-test-score-data-summary-2022.pdf

Faraj, A. K. A. (2015). Scaffolding EFL students' writing through the writing process approach. *Journal of Education and Practice*, *6*(13), 131–141.

Farjami, H. (2007). Picture production and picture perception: comparing two modes of practicing concrete vocabulary. *Iranian Journal of Applied Linguistics, 10*(1), 1–15. https://www.researchgate.net/publication/340309877_3-1_1_picture_production_and_picture_perception

Gairns, R., & Redman, S. (1999). *Working with words: A guide to teaching and learning vocabulary*. Cambridge University Press.

Hong, L. (2022), *Making a little progress your life will be different*. Global Views – Commonwealth Publishing Group.

Huang, T. J. (2005). A research on learner retention and learning support. *US-China Education Review*, *2*(1), 58–68.

Huang, Y. C. (2014). Formal vocabulary instruction in EFL context. *Journal of National Formosa University*, *32*(2), 37-50. https://doi.org/10.6425/JNHUST.201412_32(2).0004

Hulstijn, J. (1997). Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady & T. Huckin (Eds.), *Second*

language vocabulary acquisition (pp. 203–224). Cambridge University Press.

Huyen, N. T. T., & Nga, K. T. T. (2003). The effectiveness of learning vocabulary through games. *Asian EFL Journal*, *5*(4), 90–105.

Isangedighi, A. J. (1997). A comparison of study habits among the underachieving, the achieving and over-achieving junior secondary one students. *West African Journal of Education Research*, *1*, 114–119.

Karami, A., & Bowles, F. A. (2019). Which strategy promotes retention? Intentional vocabulary learning, incidental vocabulary learning, or a mixture of both? *Australian Journal of Teacher Education*, *44*(9), 25–43. https://doi.org/10.14221/ajte.2019v44.n9.2

Keshta, A. S., & Al-Faleet, F. K. (2013). The effectiveness of using puzzles in developing Palestinian tenth graders' vocabulary achievement and retention. *Humanities and Social Sciences*, *1*(1), 46–57. https://doi.org/10.11648/j.hss.20130101.16

King, K. B. (2010). *The writing template book: The Michigan guide to writing well and success on high-stakes tests*. The University of Michigan Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Kurniarahman, I. (2023). Mnemonics and their effect on students' vocabulary memorization and recall: A quantitative study. *BATARA DIDI: English Language Journal, 2*(1), 10–24. https://doi.org/10.56209/badi.v2i1.51

Le, H. S., & Trinh, M. L. (2024). An Investigation of Vocabulary Learning Strategies of ESP Students. *International Journal of TESOL & Education, 4*(1), 1–17. https://doi.org/10.54855/ijte.24411

Lestari, N., & Yulia, Y. (2018). The use of crossword puzzle to improve vocabulary mastery of the fifth grade students of SD N GOLO. *JELLT*, *2*(2), 71–79. https://doi.org/10.36597/jellt.v2i2.3275

Li, Y. (2009). *L2 learners' attitudes to English vocabulary learning strategies*. Kristianstad University.

Liao, Y. F. (2004). A survey study of Taiwan EFL freshmen's vocabulary learning strategies. *Journal of National Pingtung University, 21*, 271–288.

Lipscomb, A., Swanson, J., & West, A. (2010). Scaffolding (Ch. 21). In M. Orey (Ed.). *Emerging perspectives on learning, teaching, and technology* (pp. 226–236). https://opentextbooks.uregina.ca/teachingdiverselearners/chapter/scaffolding-2/

Mauranen, A. (2010). Features of English as a lingua franca in academia. *Helsinki English Studies*, *6(6)*, 6–28.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice*, *41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4

Merkel, W. (2016). The potential of crossword puzzles in aiding English language learners. *TESOL Journal*, *7*(4), 898–920. https://doi.org/10.1002/tesj.252

Ministry of Education. (2018). *Curriculum guidelines of 12-year basic education for elementary school, junior high and general senior high schools: Subject of English in the domain of language.* Ministry of Education.

Mustika, V., Yufrizal, H., & Huzairin. (2022). The use of crossword puzzle game to improve students' vocabulary mastery. *U-JET, 11(4),* 306–311.

Nam, J. (2010). Linking research and practice: Effective strategies for teaching vocabulary in the ESL classroom. *TESL Canada Journal*, *28*(1), 127–135. https://doi.org/10.18806/tesl.v28i1.1064

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle & Heinle Publishers.

Nemati, A. (2009). Memory vocabulary learning strategies and long-term retention. *International Journal of Vocational and Technical Education*, *1*(2), 14–24. https://doi.org/10.4314/marang.v20i1.56821

NICHD. (2000). *A closer look at the five essential components of effective reading instruction: A review of scientifically based reading research for teachers*. Learning Point Associates.

Njoroge, M. C., Ndung'u, R. W., & Gathigia, M. G. (2013). The use of crossword puzzles as a vocabulary learning strategy: A case of English as a Second Language in Kenyan secondary schools. *International Journal of Current Research*, *5*(2), 313–321.

Orawiwatnakul, W. (2013). Crossword puzzles as a learning tool for vocabulary development. *Electronic Journal of Research in Educational Psychology*, *11*(2), 413–428. https://doi.org/10.14204/ejrep.30.12186

Pillai, N. R. (2017). Using mnemonics to improve vocabulary, boost memory and enhance creativity in the ESL classroom. *Melta,* 1–25. https://www.journals.melta.org.my/TET/downloads/tet33_01_05.pdf

Puspita, N., & Sabiqoh, N. (2017). Teaching vocabulary by using crossword puzzle. *English Education: Journal Tadris Bahasa Inggris*, *10*(2), 308–325. https://dx.doi.org/10.24042/ee-jtbi.v10i2.1753

Reader's Digest. (1983). *Super word power*. Berkley Books.

Republic of China (Taiwan). (2018). *Blueprint for developing Taiwan into a bilingual nation by 2030*. Executive Yuan.

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press.

Rüdiger, S., Leimgruber, J., & Tseng, L. (2023). English in Taiwan. *English Today 153*, *39*(2). Printed in the United Kingdom. https://www.cambridge.org/core/journals/english-today/article/english-in-taiwan/DE7F4A0F324A661509D72C3A4622510B

Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge University Press.

Tran, Q. H., & Nguyen, T. M. (2022). Students' retention on online learning: Establishing a

predictive model at a private university in Vietnam. *International Journal of TESOL & Education, 2*(4), 149–172. https://doi.org/10.54855/ijte.22249

Vu, T., Winser, W., & Walsh, J. (2020). Teacher attitudes towards the English language curriculum change: The case of Vietnam. *TESOL International Journal*, *15*(6), 38–53.

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

Wei, M. (2007). An examination of vocabulary learning of college-level learners of English in China. *Asian EFL Journal*, *9*(2), 91–114.

Wilkins, D. A. (1985). *Linguistics in language teaching*. Edward Arnold.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yam News. (2023). *eNews network*. Yam News.

Yang, Y. C., & Wang, C. L. (2014). *TOEIC 900 leisure & social network vocabulary practice*. Bookman Publishers.

Yang, Y. C., & Wang, J. C. (2006). *TOEFL-iBT 120: Reading and writing.* The Jinni Publishing.

Yeh, C. Y., & Wang, Y. H. (2004). An investigation into vocabulary learning strategies used by senior high school students in Taiwan. *Taiwan Journal of TESOL*, *1*(2), 1–44.

Zitouni, M., Zemni, B., Al-Traif, H., & Aljasser, M. (2021). Vocabulary learning strategy use among translation students in Saudi Arabia. *TESOL International Journal*, *16*(4), 67–81.

Yunianta, A., Yusof, N., Othman, M. S., & Octaviani, D. (2012). Analysis and categorization of e-Learning activities based on meaningful learning characteristics. *World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences*, *6*(9), 811–816.

Biodata

Yu-Chi Yang holds a master's degree in TESOL and a Ph.D. in Adult Education. He has been teaching vocabulary-related courses at Fooyin University for 22 years. The author has published 20 books related to English proficiency exams, including TOEIC reference books, and shares his personal teaching experiences in this article.

Students' Challenges in Employing AI Tools in Legal Writing

Pham Thi Thuy Dung[1*], Luong Minh Hieu[1]

[1]Faculty of Legal Languages, Ho Chi Minh City University of Law, Vietnam
[*]Corresponding author's email: pttdung@hcmulaw.edu.vn
https://orcid.org/0009-0003-2554-7619
https://doi.org/10.54855/ijte.25524

Received: 13/08/2024 Revision: 05/03/2025 Accepted: 24/03/2025 Online: 23/04/2025

ABSTRACT

Artificial Intelligence (AI) has recently provided English learners with various interactive alternatives in classes. However, limited research has been conducted into the use of AI in English for specific purposes (ESP), especially Legal English. This paper investigates challenges students face when incorporating AI tools into Legal English Writing, stemming from either the tools themselves or their compatibility with Legal English. Forty-two undergraduate students of Legal English at the Ho Chi Minh City University of Law completed a structured questionnaire to gather quantitative data on their obstacles. The findings revealed that the primary challenges include overreliance on AI tools, educational ethics regarding plagiarism, and reduced creativity. These thoughtful insights shed light on pedagogical implications, helping guide students to use AI tools responsibly and effectively. The study also contributes to further research on AI in ESP, calling for deeper investigation to improve students' legal writing and other general purposes.

Keywords: Artificial Intelligence (AI), Artificial Intelligence Tools, Legal English, Legal Writing

Introduction

With its rapid development since 2000, English learning and teaching have witnessed substantial assistance from Artificial Intelligence (AI) with a multiplicity of AI-based systems or tools applied. Indeed, well-known book publishers like MacMillan and National Geographic Learning have introduced AI-powered adaptive learning systems that can analyze students' performance and tailor content, feedback, and practice activities to each individual's needs and learning pace. Similarly, English teachers find AI to be a helpful assistant that plays an important role in promoting their teaching efficiency. They have employed AI-based apps to help with grading and assessment, such as Grammarly, Turnitin, PlagScan, and Dupli Checker, to name but a few.

As for students, the employment of AI tools is believed to be an essential part of their learning. According to a survey of 800 U.S. college students by Pearson (2024), 56% of those who shared generative AI helped them study more efficiently, and 51% claimed it contributed to their better grades. McKinsey (2021) also found that approximately 44% of college students had used AI tools to help them complete assignments and projects. This indicates a growing trend of students incorporating AI-based tools into their learning process.

Students of legal English, similarly, have adopted the use of AI tools during their learning, especially legal writing. It has been believed that AI tools are great assistants in offering help with legal terms, grammar, and even ideas. This can be proved by the fact that students usually

CITATION | Phan T. T. D., & Luong, M. H. (2025). Students' Challenges in Employing AI Tools in Legal Writing. *International Journal of TESOL & Education*, *5*(2), 75-86. DOI: https://doi.org/10.54855/ijte.25524

do research or make corrections with the help of AI tools in their legal writing classes. However, they are not always useful for students due to some potential difficulties. Ho (2024) expressed concern that the presence of ChatGPT *(a recent AI technology chatbot)* is reshaping how university students gain knowledge and language skills, potentially disregarding the role of English language teachers. These may prevent students from utilizing AI tools in their learning process. Therefore, lecturers need to be aware of these issues so that we can have some proper interventions. In fact, it has been of great interest for researchers to investigate obstacles students might confront in their writing learning, especially academic writing. However, due to its specific features, there are few such studies on legal writing. Therefore, we, lecturers from the Faculty of Legal Languages, HCMC University of Law, desire to figure out possible challenges for Legal English students, so some suggestions could be drawn for them to fully utilize the tools in their legal writing. Accordingly, the target students in this paper are the ones with legal backgrounds, which can contribute to the significance of this study. This can be explained by Legal English's distinctive features that need careful investigation.

This paper is organized into 4 sections, with the first one introducing the study. Section 2 then offers a brief glimpse of the literature and previous relevant studies. After that, the methodology will be well demonstrated in Section 3. In detail, the writers would portray its design, respondents, and sampling procedure before describing the instrument designed. The last part of section 3 will be devoted to the description of the data collection and analysis procedure. Finally, the focus of the Section 4 is on findings and discussions. Specifically, the writer would illustrate the results and offer reasonable corresponding discussions.

Literature Review

Artificial Intelligence (AI) and AI tools

AI is formally defined in Cambridge Dictionary as "a particular computer system or machine that has some of the qualities that the human brain has, such as the ability to interpret and produce language in a way that seems human, recognize or create images, solve problems, and learn from data supplied to it ."This term is now conceived as technology enabling computers and machines to simulate human intelligence and problem-solving capabilities.

The object of this study is AI tools, which are applications run on the basis of AI and employed as assisting tools in students' writing. Some widely recognized AI tools can be listed as follows: Grammarly, Hemingway Editor, Wordtune, PaperPal, Atomic Reach, ProWritingAid, etc. In this paper, we expect to collect some fresh information about the tools favored by students of Legal English.

Challenges

In fact, challenges can be regarded in different ways. In this paper, challenges are preferred over difficulties or obstacles students encounter when utilizing AI tools to assist their legal writing.

There are some challenges regarding AI tools that are usually reported.

Challenges with ethical issues: These involve learners' struggles to avoid plagiarism, even accidentally. This has been proven to be one of the most common challenges for learners when using AI tools. Pham & Cao (2025) also agreed that on the scale of academic integrity, students often depend on AI tools to engage in dishonest practices in their learning process. (Pham & Cao, 2025)

Challenges with critical thinking skills: These are believed to ruin learners' ability to think critically. In other words, learners may have a tendency to follow the bias suggested by the tools. Challenges with autonomy are believed to cause students' lack of independence in their writing and learning generally. Yet, with special target participants and a distinguished subject involved—legal writing—this paper expects to discover challenges through different stages of AI tool application.

Legal writing

Legal writing is believed to possess some distinctive features due to its historical background in English for Specific Purposes. Traditionally, legal writing has been taught alongside legal reasoning as they are interconnected processes. There has been debate about whether to teach them together or separately. Langdell (cited in Kimball, 2006), in the late 19th century, compared the study of law to science, leading to a theoretical approach in legal education. However, this approach resulted in students' lack of proper legal writing skills. After World War II, law schools recognized the need to teach students how to write analytical legal research in plain English. Some schools combined English grammar and composition with legal research instruction. In the mid-1980s, the teaching of legal writing shifted to a process-based approach, emphasizing practice, note-taking, and correction of mistakes. Present-day legal writing pedagogy focuses on practice, legal foundations, addressing a legal audience, and the generative aspect of writing. Some professors advocate for teaching the entire sequence of thinking and developing legal arguments.

Accordingly, legal writing owns some distinguishing characteristics (Nozima, 2023).

Authority: Legal writing heavily depends on authority. Citations of authority are essential in most legal writing, as they support and validate assertions and statements made by the writer (Havard Law School Library, 2017).

Precedence: In legal writing, precedence is significant and distinct from authority. Precedence refers to the established way or manner in which legal cases have been decided in the past.

Vocabulary: Legal writing incorporates technical terminology that can be classified into four categories. First, specialized words and phrases are unique to law. Second, ordinary words have different meanings in the legal context. Third, archaic vocabulary was once common but now exists mainly in law, dating back to the 16th century. Fourth, loan words and phrases from other languages, which do not require italicization in English legal writing.

Formality: Legal writing is characterized by its formal nature, which manifests in various ways. This formality is evident in the usage of lengthy sentences, intricate sentence structures, and excessively formal vocabulary, as well as a primary emphasis on content rather than catering to the needs of readers.

Plagiarism: When lawyers write objective analyses or persuasive documents like memoranda or briefs, they are subject to the same plagiarism rules as others. However, they also face additional ethical considerations when it comes to presenting copied materials as original work (Ho, 2024).

Plain language movement: This aims to promote the use of clear and accessible language in legal documents, avoiding complex terminology and convoluted expressions. Its goal is to enhance the understandability and accessibility of legal writing.

Legalese is the use of complex and difficult-to-understand language in legal writing, making it challenging for non-experts to comprehend. This intentional obscurity excludes those without legal training and justifies high fees.

When it comes to classification, there are two broad categories of legal writing: (1) legal analysis and (2) legal drafting. The first one includes predictive analysis and persuasive analysis, which are considered to be the processes of analyzing and evaluating legal issues, cases, statutes, regulations, or legal principles in a written format. The latter refers to the creation of binding legal text. Legal drafting encompasses various types of written documents, including statutes, rules, regulations, contracts (both private and public), personal legal documents such as wills and trusts, as well as public legal documents like notices and instructions.

Due to the study's narrow scope and target participants, we are specifically addressing the second type—legal drafting—in this paper. Unlike legal analysis, legal drafting typically does not require the inclusion of legal authority citations and is typically written in a straightforward manner without a distinctive or stylized voice.

Previous Studies

Researchers have been very interested in the implications of AI tools in English teaching and learning in recent years. Plenty of papers discuss the benefits and challenges brought by AI tools in students' learning process. Researchers tend to observe the issues from two different perspectives, namely teachers' and students'.

From a teacher's perspective, Duong (2024) carried out library research on AI impacts. She expected to figure out both the positive and negative effects that AI can bring to students' academic writing at Dong Nai University. With the method of literature review employed, she concluded that with the use of AI tools, students could enjoy suggestions that were suitable for them as well as customized feedback on various domains. As for challenges, two main types were found: those with ethical dilemmas and those regarding the decline of soft skills such as critical thinking, creativity, and problem-solving skills.

With the same objectives and approach, a scoping review (2023) by a group of researchers from Indonesia also revealed similar findings. Specifically, they concluded that AI could help with individualized comments, assignments, and support thanks to its ability to meet individual learners' needs through their performance. Similar to the aforementioned study, ethical and academic integrity remained a big problem that both teachers and students faced with using AI tools.

In early this year (2024), Dugošija, a researcher from Western Serbia Academy of Applied Studies, released a paper examining the benefits and the drawbacks of implications of AI tools in English language teaching by reviewing the literature. As a result, apart from the privileges he brought, he came up with some downsides for both teachers and students. Not surprisingly, ethical issues and the lack of creativity were reported to be the challenges. Moreover, this study proved the increasing dependence upon AI, which might ruin student interaction and communication. This can be seen to be in contrast with what was stated in the previous scoping review.

Sharing the same goal and approach, Campoverde-Quezada & Valdiviezo-Ramírez published a paper in 2024 to explore the benefits and challenges of AI tools in EFL teaching and learning. As expected, the result was not so different from the others. Yet, it is worth mentioning that implementation challenges were the new issues raised uniquely in this study.

It can be easily seen that teachers explore students' challenges with AI tools in their writing thanks to previous literature. To some extent, this approach is useful for catching a brief glimpse at the issues. However, some limitations come from the lack of particularization, which might negatively influence students when AI tools are used in their writing. In fact, students may reflect on their own experiences with AI tools in different ways in accordance with the majors or the schools they belong to. Therefore, other researchers adopt the second approach based on a different outlook- the students' perception.

As a whole, these studies serve as a helpful source of references when it comes to students' challenges when it comes to the implementation of AI tools in writing. Although they could address the issue of particularizing the target students at specific levels and educational institutions, the problem with the subject involved remains, which is writing. In other words, they simply talk about EFL students. In our paper, we expect to focus on ESP students with their legal writing, which is believed to be an interesting field to examine. With that ambitious aim, our study hopes to offer a more comprehensive look at this topic.

Research Question

To fulfill the purpose of the study, the survey sought to answer the following research question:

What kinds of challenges do students face when using AI tools in their legal writing?

Methods

Pedagogical Setting & Participants

As required, students of legal English at the University of Law have to complete three modules of Legal Writing, comprising Legal Writing 1, Legal Writing 2, and Legal Writing 3, which are reasonably distributed to the school year sophomores and juniors. The modules aim to consolidate the knowledge and skills students have obtained in Legal Reading and Legal Listening. In other words, students apply fundamental knowledge to factual circumstances that give them more opportunities to get used to future working conditions. In the modules, students implement their knowledge to solve problems and improve their writing skills in legal situations, covering some areas such as the practice of law, business law, contract law, civil and criminal laws, commercial law, and so on. Then, legal students will be able to identify the key terms, types of responses, and the targets of the transactions. Also, they will be able to describe the basic features of some specific fields of law, analyze the situations, build up the relationship through letter responses, and give some legal advice to customers. These modules are taught in English. Before taking part in the research, the participants already finished the first two courses of Legal Writing. Therefore, the answers to the questions given, with a factual background of Legal English respondents, are expected to be relevant to their own experience during their implication of AI tools in their own courses.

Design of the Study

As mentioned in the first section, this paper aims to address the question: What kinds of challenges do students face when using AI tools in their legal writing? Ultimately, we expect to suggest some optimal solutions that may be helpful for both students and teachers in their learning and teaching.

With that goal aimed, the writers would love to conduct a survey with an exclusively designed questionnaire. It includes two main parts (13 questions total), with the first exploring the AI tools students employ. Then the second part comprises three sections that aim to figure out the

challenges students face regarding Accessibility, Reliability, and Reliance. Accessibility refers to the issue of whether those tools are easy to approach both mechanically and financially; Reliability is the one related to the quality of the tools; Reliance is involved with students' dependence on the use of AI tools. These three aspects of challenges correspond to three stages of AI tools employment: before, during, and after

Data collection & analysis

Due to the time limit, the convenient sampling method was adopted, and all the participants were second-year and third-year students who had finished two courses, Le, Gal Writing 1 and Legal Writing 2. In specific, the questionnaires would be delivered online to four classes majoring in Legal English, including LE46A, LE46B, LE47A, and LE47B. Then, 42 responses would be picked randomly for data analysis. As for this step, the writers would conduct an in-depth analysis to provide a factual background of the implications of AI tools by Legal English students based on their answers to the questions given.

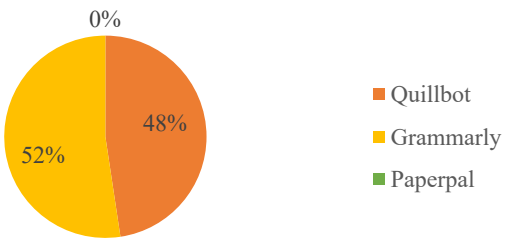
Once the findings are written, the writers would generate corresponding discussions. Initially, these are helpful for students to deal with their own obstacles. Furthermore, these useful examinations provide teachers with a heightened awareness of their students' difficulties with AI tools in Legal Writing, encouraging teachers to give assistance or conduct more effective classes.

Findings

This paper presents some of students' challenges with applying AI tools in their legal writing. It has been acknowledged that this is a single case study, so the results cannot be generalized. Nonetheless, this paper hopefully serves as a basis for future work.

Chart 1:

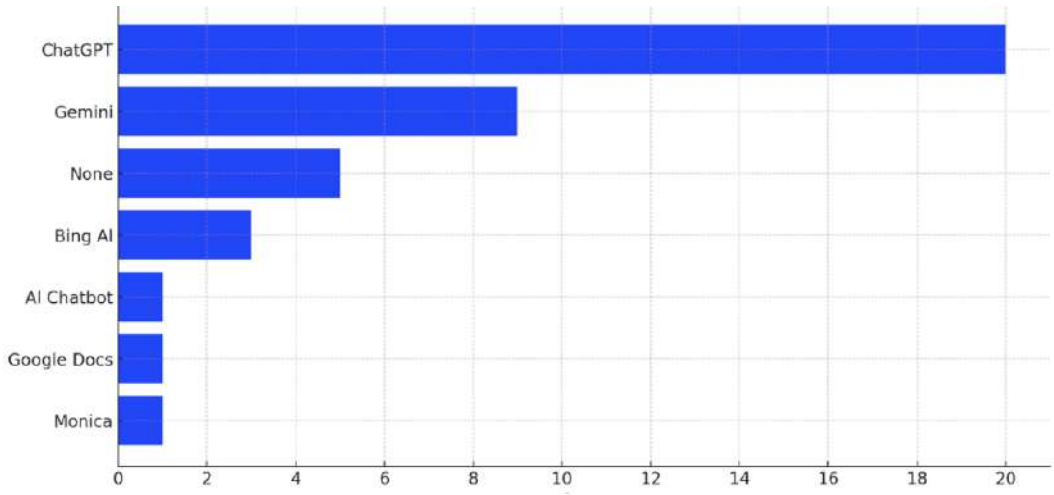
The AI Tools Employed by Students



The figure illustrates the popularity of the AI Tools used by students for Legal Writing. In particular, there was half of the students voted for Grammarly as the most well-known tool in their learning. Meanwhile, Quillbot, which was one of the most common and suitable tools on the Internet for writing learning, took over just more than 47%. The popularity of these two tools proves that students of legal English pay great attention to grammar and how they properly communicate their ideas. However, Paperpal, as suggested by the writers, happened without any selection. This means they were not getting used to this tool.

Chart 2:

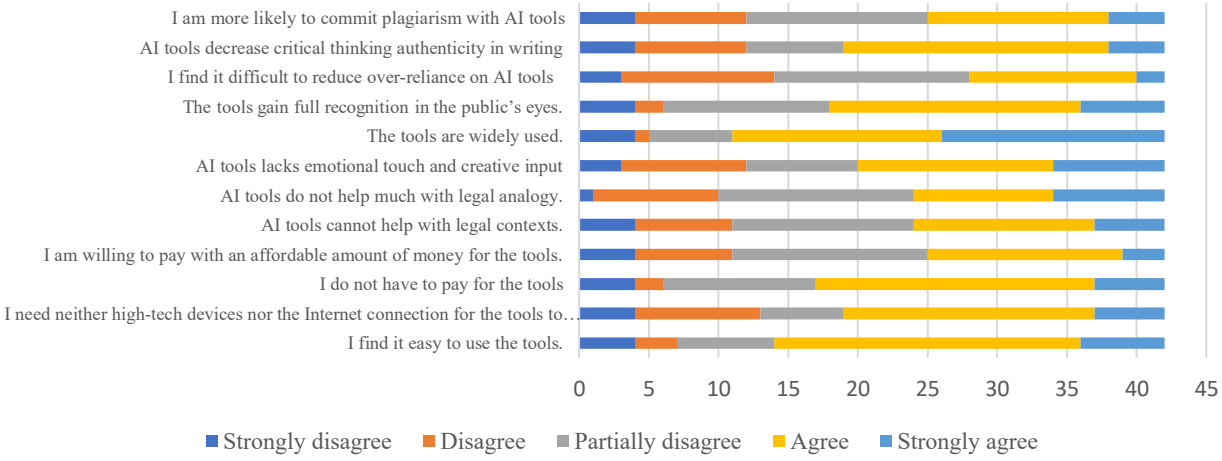
Alternative AI Tools to Help students Boost their Legal Writing Process



Besides, when requested to share the other tools, most chose Chat GPT as their alternative way to help them boost their legal writing process. Subsequently, nine (09) responses claimed the use of Gemini and three (03) choices for the tool, so-called Bing, accounting for nearly 4% and 2% of the total, respectively. More noticeably, only one (01) voted for the implementation of Google Docs and Monica in their learning process of legal writing. It can be seen that ChatGPT gained dominance over other tools.

Chart 3:

Challenges in the Employment of AI Tools



In order to discover the friendliness of the tools to the users, over 70% of participants agreed that they found it easy to use the tools for their learning, whereas some of those (only over 10%) partly disagreed with that. With regard to the feasibility of AI tools, the respondents also stated that the accessibility of these tools did not completely depend on high-tech devices as well as Internet connection. This means these tools were available and did not require complicated techniques. However, almost half of those showed their disagreements due to several technical issues. In terms of financial condition (the cost students had to pay for the tools), compared with a few responses who had to incur some expense on the full versions of the tools, many students were using the tools without payment. Besides, some of those expressed their partial

disagreement with the idea of whether they might or might not pay for the tools. More importantly, the availability of the tools experienced some differences in the willingness to pay. Specifically, the number of supporters and opponents to this criterion was similar, with an average of 40% of the total. However, there were some answers, which were "strongly disagree". Evidently, there would be great concern if they had to pay to access the tools on their own. Therefore, we can conclude that most students do not find it difficult to get access to AI tools in technical terms. Yet, that can be problematic if the tools require full purchase.

Another practical criterion was the quality of the tools. In other words, the participants had to determine the tools' reliability in legal contexts. Students appear to be not sure about whether the tools can deal with legal contexts so that the results offered are good enough. This indeed can be seen by the same number of those who agree and partially agree (13 students for each). However, students tend to put their trust in AI tools' expertise in legal words. Only 18 students doubted the efficiency of the tools when they had to deal with legal terms. Surprisingly, not all of the participants think that the tools will have trouble with emotional touch and creative input. Half of the participants didn't make up their minds to give a judgment on the affective aspect that AI tools can provide them with. At the same time, the other half holds a strong belief that the tools will fail to produce and process the information with emotions and creativity. The result reveals an interesting fact that learners do not impose grave doubts on the tools when dealing with legal jargon or affective aspects in legal writing.

The answers to the last two questions related to reliability reveal a very interesting tendency of AI tool users. While a majority of students (more than 30) admit that their tools are very popular, not all use highly recognized tools. Around 10 students are reluctant to state if the tools they employ are well estimated. This can be interpreted as learners not having full consciousness of the tools' reliability.

Meanwhile, when being asked about their dependence on AI tools, more than 50 percent of participants are aware that the tools worsen their critical thinking in writing. Nevertheless, around 15 of them find it challenging to get over their reliance on the tools while the others don't. This might be in relation to their responses to the last question about the risk of committing plagiarism. About 15 students admit that they are more likely to get into plagiarism with the use of AI tools. The figure means learners do not find it a real challenge to get over the dependence on the tools, although their critical thinking skills might be affected. Unfortunately, this signals potential reliance on learners' highly frequent use of AI tools.

Discussion

Data gathered from the survey has indicated that students are using more than one type of AI tool over their study time. Subsequently, they chose Quillbot and Grammarly as their popular legal writing assistants. Chat GPT is also their great choice. The figures collected reveal that these tools are highly intuitive and readily accessible to them at any time. They agreed to adopt these advanced technological solutions to justify their choices since the requirements are less complicated, and the users do not need more mechanical techniques to operate these tools. In particular, it is feasible for students to access these without the need for an Internet connection or any high-tech devices. Indeed, these tools are always at the top of the teaching and learning assistance list. They are available to the public and approachable at any time. More interestingly, students allow themselves to need more help from other tools such as Gemini, Bing, Google Docs, or Monica.

Aligning with previous research findings, this paper's findings shed light on several noticeable challenges and dilemmas that students might encounter when using AI Tools. These challenges can be categorized into three areas: (1) learning autonomy and over-reliance, (2) educational ethics, and (3) lack of creativity.

Lack of Learning Autonomy

Students obviously acknowledge one of the significant issues when using AI tools in their learning in these courses is that they cannot increase their autonomy. In fact, AI tools help them with thorough ideas, well-structured outlines, and accurate sentence structures in less amount of time. Therefore, students find it hard to control the abuse of these tools in their learning process. Meanwhile, it comes to a statement that "the learner's psychological relation to the process and content of learning - a capacity for detachment, critical reflection, decision-making, and independent action" (Little, 1991, as cited in Morbedadze, 2015, p.2). Duong (2024) agreed that depending on AI tools for structures and feedback potentially reduced their (students') critical thinking, creativity, and problem-solving abilities. Therefore, Dugošija's (2024) concluded that an abuse of AI tools in learning can deter students' ability to interact with the material meaningfully. In other words, due to a lack of learning autonomy, in the courses of legal writing, where independent reasoning and argumentation are crucial, students are in trouble producing their own language as well as developing the analytical skills needed in legal discourse. However, it is also evident that students showed their hesitation over whether they can reduce over-reliance on AI tools. As mentioned before, this turns out to be a signal for a kind of addiction that learners cannot control. Unlike previous studies on students' challenges with AI tools, the result of this dire prediction about the lack of learning autonomy leaves an issue for teachers and students to consider when using AI tools.

Educational Ethics

Dam (1990, as cited in Gathercole, 1990) agreed that learning autonomy is the willingness and capacity of students to control and oversee their own learning. Therefore, a lack of learning autonomy may lead to an ethical issue in relation to plagiarism. Indeed, students tend to use content generated by AI tools to blend with their own works without valuing the precise expressions, along with the significance of authenticity in legal writing. Unfortunately, students do not express their great concerns about this problem. This finding is similar to what was found in the studies by Duong (2024) and Dugošija (2024). Moreover, most of the students are reported to be willing to spend on the tools if they are required to pay due to their high estimation of the tools' efficiency. Therefore, it is essential for teachers to develop clear guidelines as well as adequate assistance to prevent learners' overuse of AI tools in their writing learning.

Limitations of Creativity

AI Tools can place learners in legal contexts, where students can understand the situations and implement the key terms in specific cases. However, many of them are worried about creative input. Malik et al. (2023) addressed similar findings that AI-generated works could hinder students' creativity, which might risk their learning process of legal writing. Instead of creating ideas and constructing persuasive arguments, students depend on the available sources or the contents produced by the tools. With assistance with grammar, structures, and vocabulary, the contents may lack the individualized voice and personality. Dugošija (2024) noted that responses provided by AI tools might diminish students' creativity. Consequently, students claimed the lack of emotional engagement in learning legal writing.

Conclusion

To fulfill the aim of legal writing lessons, the application of technological advances, particularly AI Tools, has become essential. It is worth affirming the benefits of the tools that can be brought into the students' process of legal writing. However, it should be noted that the abuse of AI Tools causes some challenges. The use of these tools requires students to consider whether the contents they are using in their situations are reasonable and accurate. Moreover, students' reliance on AI tools leads to the ruin of critical thinking. Moreover, potential plagiarism can be a major issue that all language learners must be fully aware of.

With the paper's findings, we can make some suggestions for teachers and learners on how to fully utilize AI tools in their legal writing and minimize their undesirable effects. First and foremost, educational institutions can consider offering support for learners who struggle to get full access to the tools, which might hinder them from considerable assistance for their writing learning. The support can be realized by the establishment of a laboratory or a library where learners can come and do research in which they need help from fully equipped AI tools. This might be troublesome in some cases when learners have to be at school for any search or investigation. Yet, this turns out to be helpful for both teachers and learners to control their unconscious addiction or reliance on the tools. In fact, they have time for their creativity and critical thinking to develop rather than spending whole days with their AI assistants.

Additionally, it is essential to measure the reliability or the efficiency of the tools used. Teachers are believed to take responsibility for providing their students with practical guidelines on how to make the best use of AI tools. In order for this to be done, educational institutions can hold some training or talk shows on this issue. As a result, learners might have better choices in AI tools with greater reliability, which brings about better assistance for their writing learning.

As for the fear of learners' overreliance on AI tools, it is suggested that teachers and schools raise learners' awareness of their potential negative effects. Apart from the training and talk shows mentioned earlier, teachers need to be strict when dealing with ethical issues or lack of autonomy caused by overdependence on the tools. That is necessary for learners to have a full understanding of when and how to use the tools properly.

Acknowledgments

We extend our high appreciation to the Faculty of Legal Languages for their exceptional teaching and support and to the legal students who made invaluable contributions and were willing to join this research paper to share their experiences using AI tools in their learning process of legal writing. Their insights and encouragement have been instrumental in the completion of this study. Lastly, we are grateful to Van Lang University for providing resources and facilities that support our research endeavors.

References

Alkamel, M. A., & Alwagieh, N. A. (2024). Utilizing an adaptable artificial intelligence writing tool (ChatGPT) to enhance academic writing skills among Yemeni university EFL students. *Social Sciences & Humanities Open, 10*, 1-9, https://doi.org/10.1016/j.ssaho.2024.101095.

Campoverde-Quezada, D. A.-R. (2024). The Double-Edged Sword: Benefits and Challenges that Artificial Intelligence Tools can Bring to EFL Teaching and Learning. *Revista Metropolitana de Ciencias Aplicadas, 7*(2), 304-316.

Cummings, R. E., Monroe, S., & Watkins, M. (2024). Generative AI in first-year writing: An early analysis of affordances, limitations, and a framework for the future. *Computers and composition, 71*(2), 1-11, https://doi.org/10.1016/j.compcom.2024.102827.

Dugosija, T. (2024). Benefits and challenges of artificial intelligence in English language teaching. *Knowledge International Journal, 62*(2), 275-280. Retrieved from https://ikm.mk/ojs/index.php/kij/article/view/6590.

Ho, P. P. (2024). Using ChatGPT in English Language Learning: A Study on I.T. Students'Attitudes, Habits, and Perceptions. *International Journal of TESOL & Education, 4*(1), 55-68. DOI: https://doi.org/10.54855/ijte.24414

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective. *International Journal of Eduational Research Open, 5*, 1-11, https://doi.org/10.1016/j.ijedro.2023.100296.

McKinsey. (2021). *The state of AI in 2021.* McKinsey & Company.

Nozima, A. (2023). The Distinguishing Features of Legal Writing in English. *International Journal of Formal Education, 2*(5), 119-124. Retrieved from https://journals.academiczone.net/index.php/ijfe/article/view/819/707.

Pham, T. M., & Cao, T. T. (2025). The Practice of ChatGPT in English Teaching and Learning in Vietnam: A Systematic Review. *International Journal of TESOL & Education, 5*(1), 50-70. DOI: https://doi.org/10.54855/ijte.25513

Shen, X., & Teng, M. F. (2024). Three-wave cross-lagged model on the correlations between critical thinking skills, self-directed learning competency and AI-assisted writing. *Thinking Skills and Creativity, 52*, 1-13, https://doi.org/10.1016/j.tsc.2024.101524.

Shofiah, N., Putera, Z. F., & Solichah, N. (2023). Challenges and opportunities in the use of artificial intelligence in education for academic writing : A scoping review. *Conference Psychology and Flourishing Humanity (PFH 2023)*, 174-193, https://doi.org/10.2991/978-2-38476-188-3_20). Atlantis Press.

Stornaiuolo, A., Higgs, J., Jawale, O., & Martin, R. M. (2024). Digital writing with AI platforms: the role of fun with/in generative AI. *English Teaching: Practice & Critique, 23*(1), 83-103, https://doi.org/10.1108/ETPC-08-2023-0103.

Strickland, C. J. (2012). The Dark Side of Unattributed Copying and the Ethical Implications of Plagiarism in the Legal Profession. *North Carolina Law Review, 90*(3), 919-956. Retrieved from https://scholarship.law.unc.edu/nclr/vol90/iss3/9/

Biodata

Pham Thi Thuy Dung holds a master's degree in TESOL (by University of Social Sciences and Humanities). She currently works as a lecturer at Faculty of Legal Languages, Ho Chi Minh City University of Law. Her research interests range from General English teaching as well as Legal English teaching methodology.

Luong Minh Hieu holds a master's degree in education with a specialization in TESOL

(conferred by the University of Southern Queensland, Australia). He is currently working as a full-time lecturer at Ho Chi Minh City University of Law, and has been teaching legal English to both English-major and non-English major students. In addition, his research interest lies in the fields of Applied Linguistics, Teaching Methodology, and Legal English.

Appendices

1. What AI tools do you have a chance to employ during your Legal writing?
   a. Quilbot
   b. Grammarly
   c. Paperpal
   d. Other(s):
2. Put a tick at the columns that suit you the most

	Strongly disagree	Disagree	Partially disagree	Agree	Strongly agree
A. Accessibility					
I find it easy to use the tools.					
I need neither high-tech devices nor the Internet connection for the tools to be used.					
I do not have to pay for the tools					
I am willing to pay with an affordable amount of money for the tools.					
B. Reliability					
AI tools cannot help with legal contexts.					
AI tools do not help much with legal analogy.					
AI tools lacks emotional touch and creative input					
The tools are widely used.					
The tools gain full recognition in the public's eyes.					
C. Reliance					
I find it difficult to reduce over-reliance on AI tools					
AI tools decrease critical thinking authenticity in writing					
I am more likely to commit plagiarism with AI tools					

Students' Perceptions of AI Language Models as Virtual Assistants in Learning Writing- A Case Study at a Tertiary Institution

Pham Ngoc Thai Binh[1], Tran Thi Mai[1*]

[1]Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam
[*]Corresponding author's email: mai.tt@vlu.edu.vn
[*]https://orcid.org/0000-0002-5785-810X
https://doi.org/10.54855/ijte.25525

Received: 26/07/2024 Revision: 24/04/2025 Accepted: 29/04/2025 Online: 03/05/2025

ABSTRACT

Research trends in AI language models for writing assistance are increasing, yet a gap exists concerning their impact on language learning, especially in Vietnam, as well as stakeholder attitudes. This study explores perceptions of 147 students, mainly English majors at Van Lang University, regarding AI tools such as ChatGPT, Poe, and Gemini in English writing. Combining quantitative and qualitative analysis through the use of questionnaires and in-depth interviews, the research found a generally favorable student attitude, recognizing AI's accessibility and flexibility. However, concerns about over-reliance and accuracy were identified. Highlighting implications for effective AI integration in Vietnamese writing pedagogy, this study contributes to understanding student perspectives on this evolving educational technology.

Keywords: AI language models, virtual assistant, learning writing

Introduction

In recent years, the integration of AI technology into education has grown significantly due to its numerous advantages in enhancing teaching methodologies and classroom management (Chassignol et al., 2018). Specifically, AI language models have revolutionized language learning by providing automated feedback and personalized writing assistance, and real-time grammar and style suggestions. These innovations have made AI tools increasingly popular as they help students improve coherence, accuracy, and fluency in writing. Moreover, AI-powered virtual assistants offer interactive learning experiences, enabling learners to refine their writing skills through instant corrections and adaptive feedback. By streamlining the writing process, AI not only supports students in developing their writing proficiency but also fosters independent learning and critical thinking (ALAfnan et al., 2023).

Traditional methods of teaching writing, such as classroom instruction, peer assessments, and teacher feedback, have long played a crucial role in developing students' writing skills. These traditional approaches are now complemented by AI-powered virtual assistants that provide instant feedback, correct grammar errors, suggest improvements and facilitate collaborative writing. Such applications as ChatGPT, Grammarly and ProWriting Aid enjoy widespread utilization across educational, professional, and daily contexts since they offer automated

CITATION | Pham, N. T. B., & Tran, T. M. (2025). Students' Perceptions of AI Language Models as Virtual Assistants in Learning Writing- A Case Study at a Tertiary Institution. *International Journal of TESOL & Education*, *5*(2), 87-115. DOI: https://doi.org/10.54855/ijte.25525

written corrective feedback utilizing technology to anticipate subsequent sentences or words in a dialogue or textual command (Fitria, 2023). In simpler terms, they function as chatbots – computer programs designed as virtual robots capable of simulation human-like conversations.

From teachers' perspectives, AI applications can be beneficial in student writing, specifically content quality and organization (Marzuki et al., 2023). They, furthermore, can boost students' engagement and motivation thanks to their natural language capabilities (Baskara, 2023). Not only are these applications studied through descriptive research, but there have been empirical works, such as AlAfnan et al. (2023), showing that ChatGPT, a notable AI chatbot, benefited student writing in various ways, from providing input high in accuracy and reliability, to being a platform for students to look for ideas for theory-based and application-based problems.

As AI language models evolve in sophistication and AI tools become increasingly integrated into education, it is essential to examine student's perceptions and interactions with these technologies in writing instruction. This is because understanding their experiences can help enhance pedagogical strategies and effectively incorporate AI into the writing learning process (Micheni et al., 2024). In the Vietnamese context, there have been a few research studies looking into the roles of these AI-powered writing assistants. However, most of the studies explore the topic from the teacher's views. Among a few that seek student's perceptions, they lack a theoretical framework (Nguyen et al., 2023; Tran, 2024).

By delving into students' attitudes, interests, and the challenges they might face, this study aims to contribute to this ongoing discussion, providing a more thorough understanding of students' perspectives on the effectiveness of AI tools in the learning-to-write process. In particular, the conclusions of this research may be useful to both the students and the faculty members at tertiary education level. This study extends its applicability, employing AI support for learning writing, as well as those utilizing AI for broader educational purposes, in particular, who may experience and recognize themselves in this research. This research could be referenced by the instructors to gain a deeper understanding of students' utilization of virtual tools. To the greatest possible extent, the standard of education at the university involved in the research was conducted will have some proposals related to combining AI with education in the coming years.

Literature Review

Theoretical Framework

This study is based on the Technology Acceptance Model (TAM). Originally developed by Davis (1989), TAM explains technology acceptance through two key factors: Perceived Usefulness (PU) and Perceived Ease of Use (PEU), depicting how users adopt and utilize a particular technology. TAM, therefore, can serve as an appropriate framework for this study because it provides a structured way to examine how students adopt and perceive AI-powered writing tools.

In the context of this study, PU refers to students' beliefs that AI writing tools can improve their writing proficiency, such as by offering grammar correction, instant feedback, and content suggestions. Meanwhile, PEU reflects whether students find these tools easy to use and integrate into their writing learning process. If students perceive AI as both useful and effortless to use, they are more likely to accept and rely on it as a writing assistant.

TAM has been widely applied in educational technology research (Salloum et al., 2019; Venkatesh & Bala, 2008), particularly in studies analyzing user adoption of digital learning

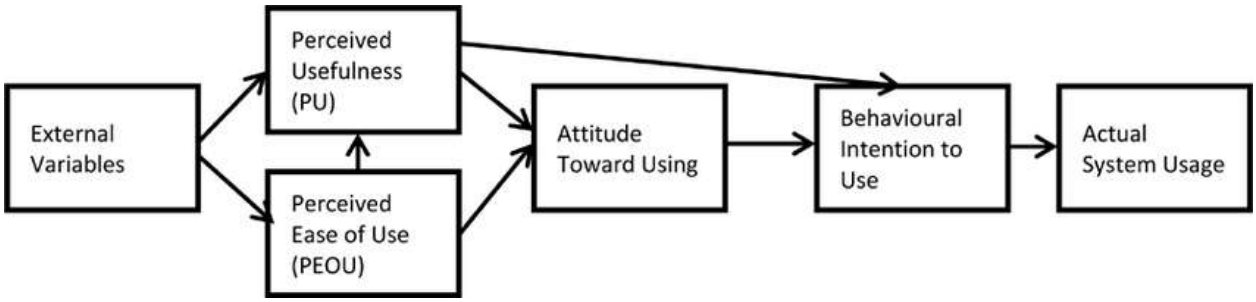
tools. Given the increasing presence of AI-powered writing assistants, TAM provides a clear framework for understanding how students interact with these tools and what factors influence their attitudes toward AI-assisted learning.

Additionally, this study seeks to fill a research gap in the Vietnamese context, where limited studies have explored students' perceptions of AI writing tools. Using TAM allows for a systematic analysis of whether Vietnamese students find these tools beneficial and easy to use, as well as any challenges they face in adoption. Furthermore, by examining students' attitudes, this research can generate practical insights for educators, AI developers, and policymakers. Educators can use these findings to refine how AI is integrated into writing instruction, while developers can leverage the results to enhance AI tools to better meet students' needs.

Overall, TAM provides a solid theoretical foundation for this study by helping explain why students might adopt AI in their writing and what factors influence their acceptance or hesitation. By applying this model, the research contributes valuable knowledge to the ongoing discussion on AI's role in language learning and its potential impact on writing education.

Figure 1.

Framework of Technology Acceptance Model (TAM)



AI Language Models

Large Language Models

Large language models (LLMs) are smart computer programs that understand and create human-like language. According to Floridi's fifth concept (2023), LLMs are revolutionary due to the successful management of the electromagnetic qualities to analyze text and frequently give results indistinguishable from those humans can create. These models have many parameters that conduct pre-training tasks (such as implicit language modeling and autoregressive prediction) to recognize and interpret human language by modeling contextual and probabilistic text semantics from vast amounts of text data (Yang et al., 2023). They can learn word collocations, grammar rules, and the meaning behind sentences. Therefore, Entering the LLMs field is a supercharged language tool, and they are impacting language teaching and learning today (Roose, 2023).

The discussion of LLMs inevitably leads to the foundational field of Natural Language Processing (NLP). The historical development of NLP can be divided into three periods. In the 1950s, Alan Turing introduced the "Turing Test", marking the beginning of NLP ideas. From the 1950s to the 1970s, the rule-based era attempted to replicate the way humans understand language using shared rules. There were substantial improvements in Internet technology between the 1970s and the early 2000s. Researchers have used statistical methods, simplifying NLP problems into probabilistic problems and making breakthroughs in tasks such as language translation and text classification (Li, 2024).

Due to the progress made in the fields of AI and NLP, there has been the emergence of intelligent

tutoring systems and adaptive language learning platforms. One such tool is ChatGPT, developed by OpenAI, which utilizes large-scale language models to generate human-like text based on a given input (Bender et al., 2021). It comes in two versions: a free one is ChatGPT 3.5, and a paid one is GPT-4. OpenAI likely uses feedback from the free version to improve the paid one. There are other platforms, such as Microsoft's GitHub Copilot (a tool developed in collaboration with OpenAI) and Gemini (previously Bard) by Google (Tira, 2023).

It is clear that these LLMs benefit students, teachers, and researchers, especially non-native English speakers, as they enhance academic writing quality. These models can interpret natural human input by drawing from the most appropriate human-requested topics. They also create writing templates derived from the requested content, including unique creativity, interaction styles, and work culture (Bonner et al., 2023).

Neural Machine Translation

Neutral machine translation (NMT) is the main method in systems aimed at translating natural language sentences using actual computers, relying on previous methods based on linguistic rules and knowledge (Tan et al., 2020). It adopts deep learning algorithms to provide more accurate and consistent translations, trained on large bilingual linguistic corpora, aiming to provide reliable translations based on high-probability sentences (Linguaserve, 2023). Instead of relying on many fine-tuned sub-components, NMT attempts to build and train a single, large neural network capable of understanding input sentences and producing accurate translations. It seeks to optimize translation performance through a unified neural network, unlike statistical machine translation, which relies on discrete components (Bahdanau et al., 2014).

Popular online translation platforms such as Google Translate and Microsoft Bing Translator have consistently progressed and integrated the latest developments in NMT to enhance the quality of translation outcomes. The evolution of NMT has transformed the field of machine translation, making it more accessible, accurate, and adaptable to a wide range of languages and contexts.

AI Language Models in Language Learning

The utilization of large AI language models in language classrooms is widespread, with applications spanning all educational levels – from primary and secondary to tertiary and professional development. According to Kasneci et al. (2023), these models contribute significantly to enriching the learning and teaching experiences. They help elementary school students develop reading and writing skills, critical thinking, and comprehension by suggesting corrections, generating questions, and explaining complex texts. In middle and high school education, they aid in language acquisition and master diverse writing styles by generating practice problems and quizzes. University students benefit from these models in research, writing tasks, and critical thinking processes by providing summaries, outlines, and curated resources on specific topics, enhancing their understanding and analytical capabilities when engaging with academic material. Considering that advancements in artificial intelligence are bound to result in continuous enhancements in various language-related activities, it might be opportune to accept the fact that students will increasingly have access to sophisticated writing assistance. In response, there is a need to explore strategies for offering suitable guidance in this context (Carvalho et al., 2022).

In language classrooms, AI can act as a language teacher, promoting personalized learning. Previous studies have shown that AI offers advantages, the first of which is the ability to use chatbots anytime, anywhere, and increased confidence in learning activities compared to human tutors (Sumakul et al., 2022). Secondly, the AI chatbots enable students to participate in

conversations and to receive instant feedback on their language usage and understanding. AI chatbots give students, especially introverted ones, the opportunity to ask questions that they might not feel comfortable asking teachers or peers in class. This creates a low-pressure environment for students to enhance their language skills through practice (Amin, 2023). Likewise, Xiao and Zhi (2023) stated that students see ChatGPT as a learning companion or personal tutor, delivering personalized, readily accessible, and adaptive feedback. However, a number of students remain critical of AI's suggestions and feedback. They demonstrate the capacity to think critically about the information generated by ChatGPT. They reported to modify prompts, train the model, verify and selectively accept the information provided. (Xiao and Zhi, 2023).

The findings of the reviewed studies offer practical insights for using ChatGPT in English teaching and learning across countries. Most of these studies provide positive outcomes of this tool. However, Kasneci et al. (2023) and Xiao and Zhi (2023) have mentioned issues related to plagiarism, copyright, and understanding the source language when utilizing ChatGPT. Plagiarism concerns arise from the ease with which students can generate AI text and submit it as their own, while copyright challenges stem from uncertainties about ownership of AI-created content. The researchers have suggested the feasibility of integrating tech-advanced tools into language classrooms, highlighting the importance of offering guidance to help students use ChatGPT appropriately and effectively. However, since ChatGPT is not the only AI tool available that learners can access, additional research on other LLMs is required to be conducted to discover any other potentials and challenges in greater depth, allowing better utilization of these tools in writing instructions.

Benefits of AI-Language Models as Writing Assistants

In the realm of second-language writing, there are unique challenges stemming from potential shortcomings in lexical, syntactic, pragmatic, or rhetorical knowledge. Navigating through these challenges to offer corrective feedback that genuinely benefits writers proves to be an impressive undertaking (Godwin, 2022).

Notably, ChatGPT, Grammarly, and Google Translate have emerged as advanced tools commonly used to assist in the face of these complexities. These text-based AI models are often freely available on the Internet, and premium versions are available that come with extra features. Advanced systems, such as GPT-3, offer complete texts that require only a general topic or prompt to function effectively (Alharbi, 2023). They can provide detailed suggestions and recommendations for the adaptation of text elements within seconds.

ChatGPT is a versatile tool that supports language learning in various ways, including text generation, translation, summarization, problem-solving, content and grammar correction, vocabulary enhancement, and question-answering. Intelligent writing assistance tools have also improved by generating texts spontaneously and independently, providing linguistically acceptable suggestions and improvements for word choices. With these features, ChatGPT aids in the improvement in students' writing (Athanassopoulos et al., 2023). Among other tools, Google Translate specializes in text translation. Meanwhile, Grammarly is for improving spelling, analyzing text, rewriting, summarizing, and checking for plagiarism (Jumriah et al., 2024). Also, these AI tools are considered useful in assisting English language learners to prepare for such standardized tests, advancing demorcatising global language test preparation (Sari, 2024).

In the Vietnamese context, Ho (2024), Ha and Ho (2025), Nguyen (2023) concur that ChatGPT can provide students with instant and personalized feedback for their writing, and students can

revise their works, either essays or academic papers, from ChatGPT's suggestions. This aids students in enhancing multiple aspects of their writing skills, from lexical resources and accuracy, grammatical range and accuracy, to the organization and development of their ideas. Furthermore, students' motivation in their writing training can be boosted thanks to the benefits ChatGPT can offer (Tran, 2024). Similarly, in an experimental study of Chu et al. (2024), AI writing tools, including ChatGPT and Grammarly, are proved to foster the engagement of students in their English skills, including writing skills. Meanwhile, Poe, another AI-power writing assistant, is revealed to stimulate tertiary students' interest in one aspect of learning writing, which is vocabulary (Pham et al., 2024).

Disadvantages of AI Language Models as Writing Assistants

Despite their potential benefits, there is an ongoing debate about the impact of AI language models on education, with concerns raised about their limitations and ethical issues.

Firstly, there are arguments about the quality of AI-generated feedback. Such tools as Grammarly are reported to perform quickly in correcting grammatical errors, yet fail to provide in-depth comments on the logic, creativity, and style in learner's writing (Duong & Le, 2024; Ha & Ho, 2025), misleading the students into thinking that improving writing work is simply correcting grammatical mistakes, leaving out coherence and cohesion.

In a worse scenario, LLMs such as ChatGPT and Google's Gemini are even reported to have a phenomenon called "hallucination", meaning their content can include fabricated information presented in the form of facts. This is caused by source- reference divergent data used to train the model. The problem can also root in the modelling process when there are errors in text encoding and decoding (Athaluri, 2023; Lee et al., 2022; Tira, 2023; Ziwei et al., 2023). Phan (2023) also points out this problem of AI writing assistants, discovering sometimes they provide students with inaccurate responses. However, the study does not dive into how critical students are when encountering the problem.

Furthermore, the overuse of AI-powered tools can lead to laziness and reduce students' ability to absorb and interpret language. Tira (2023) has mentioned that modern technology is threatening education as cheating in stealing ideas and plagiarism becomes easier. For instance, ChatGPT has the ability to create English essays, which raises ethical concerns that students can use these tools to cheat. This ease of generating text can foster laziness by reducing the cognitive effort students spend in actively engaging with the language, such as choosing words and constructing sentences themselves. Moreover, when students bypass active language processing through AI assistance, they may fail to encode new vocabulary, grammatical patterns, and contextual nuances, ultimately hindering their ability to absorb and interpret the language effectively. The existence of laziness and cheating will continue without the educational values of integrity and respect. The existence of laziness and cheating will continue without the educational values of integrity and respect. It is important to remember that technology was created to be an assistant, not a scammer, yet what is alarming is that these chatbots lack contextual comprehension, critical thinking abilities, and the capacity to make ethical decisions (Tira, 2023).

Relevant training is essential to enhance users' digital literacy skills and prevent misuse. The lack of detailed feedback from ChatGPT and the need for human supervision, especially teachers' instruction and guidance, are also subjects of concern (Athanassopoulos et al., 2023; Ho, 2024).

Students' Perceptions of Using AI-Language Models

As artificial intelligence (AI) tools become increasingly integrated into writing instruction, researchers have explored students' perceptions of their benefits and limitations, revealing both positive attitudes and significant concerns regarding their effectiveness and ethical implications.

Sumakul et al. (2022) revealed that EFL learners generally hold favorable attitudes toward AI technology in their writing classes. These positive perceptions stem from AI's ability to enhance comprehension of theoretical concepts, facilitate the writing process, and support grammar and vocabulary acquisition. However, students perceive AI tools as less effective for tasks requiring text analysis or summarization.

Despite these advantages, concerns regarding the limited accuracy of AI-powered writing tools are widespread. Phan (2023) found that tools such as Google Translate and Grammarly occasionally produce inaccurate outputs, raising doubts about their reliability. Additionally, some students express apprehension about AI's potential impact on their writing development, fearing that over-reliance on these tools may reduce their motivation to learn, hinder independent writing skills, or erode personal writing styles. The risk of creating dependency on AI-generated suggestions is another frequently cited drawback.

Beyond academic concerns, students also show their fear of ethical and professional apprehensions. Some of them are afraid that AI advancements could lead to job displacement, particularly in fields such as teaching and translation. Moreover, issues of data security and privacy breaches, including risks of data leaks, surveillance, or unauthorized access to personal information, have emerged as critical concerns. Plagiarism and fairness issues are also debated; while AI tools provide comprehensive writing support, some students argue that they create unfair advantages, particularly when premium features are restricted to paid subscribers (Burkhard, 2022).

The studies discussed above employed both qualitative and quantitative research methods to investigate students' perceptions of AI-assisted writing. While most findings indicate positive attitudes toward AI, the effectiveness of these tools largely depends on students' perspectives and evolving learning processes. Moreover, quantitative studies in the Vietnamese context about the voices of students themselves are limited in number and rely on small sample sizes as well as short intervention periods, which may restrict the generalizability of their findings. Given the rapid advancement of AI language models, there is an urgent need for more research, particularly in second-language learning settings, to provide deeper insights into AI's long-term impact on students' writing skills and academic development.

Research Questions

The following are the research questions that were developed based on the purpose of this investigation:

- Research question 1: What are students' attitudes towards using AI language models as English writing assistants?
- Research question 2: How do students perceive the effectiveness of AI language models as virtual assistants in improving their writing skills?
- Research question 3: What challenges do students face in using AI language models in learning writing?

Methods

Pedagogical Setting & Participants

This study was conducted at Van Lang University in Ho Chi Minh City, with participants drawn from various disciplines, primarily English Language majors who frequently engage in writing tasks. Their English proficiency levels range from basic to advanced as they pursue professional certifications in their respective fields.

At Van Lang University, writing instruction varies by discipline. For English Language majors, it is taught as a standalone module, while for students in other disciplines, it is integrated with other language skills. In writing classes, students combine traditional learning methods with AI language models for various purposes, including grammar correction, text translation, and paraphrasing, to enhance the clarity and readability of their essays.

As for the quantitative data, convenience sampling was utilized for the survey portion. The researcher targeted classes at Van Lang University, specifically Writing and other specialized English courses where students have a strong foundation in writing skills and frequent practice. This approach offered practicality in terms of time, effort, and cost-effectiveness (Acharya et al., 2013; Golzar et al., 2022). A total of 147 students from various majors participated in the survey. Notably, the year-level distribution showed a higher concentration of final-year students (72.8%), followed by third-year (14%), second-year (11%), and first-year students (2.2%). All participants were actively engaged in developing their English writing skills, and a majority reported using AI language models. These data were shown in the demographic questions in the questionnaire.

Regarding qualitative data, to gain deeper insights, the study additionally employed voluntary sampling. Ten students majoring in English Language from Van Lang University volunteered for semi-structured interviews, providing valuable qualitative data on the research topic.

Research Design

A mixed-methods approach was employed in this study, integrating both qualitative and quantitative methods to enhance the accuracy and reliability of the findings. These two methodologies are complementary and mutually reinforcing, allowing for a more comprehensive understanding of the research problem. As Kelle (2006) emphasizes, combining qualitative and quantitative approaches, especially in social science, enables researchers to capture both depth and breadth in their analysis. Similarly, Fraenkel et al. (2012) and Almalki (2016) both argue that these methods work in synergy, each contributing unique strengths to ensure more convincing and well-rounded results.

The quantitative design is to produce a questionnaire with a 5-point Likert scale on which participants can score their level from 1 (strongly disagree), 2 (disagree), 3 (no opinion), 4 (agree), and 5 (strongly disagree) with the suggested statement. The researcher chose a 5-point Likert scale instead of a 4-point Likert scale to avoid lacking neutral opinions from respondents. The 4-point Likert scale has lower accuracy and provides less ability to clarify a point of view (Wittink & Bayer, 2003). Meanwhile, the study of Joshi et al. (2015) revealed that the 5-point Likert scale has been recommended by most of the researchers because it will reduce the frustration of respondents' patience and the rate and quality of response will be increased. It is applied to find statistics, predict, discover cause-and-effect relationships, and apply knowledge to larger participant groups.

The semi-structured interview method is used to design qualitative research. This includes asking a basic set of questions, but it also allows conversations to deepen and develop in ways

that participants actively address. This method aims to collect high-quality and insightful information on the study topic. According to Bhandari (2020), the qualitative method focuses primarily on language, and it is used to understand diverse ideas, beliefs, and perspectives.

In summary, the mixed-method approach assists the researcher in answering the three questions in this study since the 5-point Likert scale may be examined at various levels, including neutral viewpoints, while interviews will directly clarify students' awareness of employing AI language models in their writing learning. Combining both methods in a mixed approach will supply findings from general to specific, helping readers better understand the study.

Research Instruments

This study has applied a quantitative design approach using a questionnaire. This survey consists of 33 items in total and is separated into four sections: students' attitudes, students' perceptions, benefits, and drawbacks of AI language models.

Table 1

The Number of Items, Cronbach's Alpha, and Example Item

Criteria	Number of items	Cronbach's alpha	Example item
Students' PU	7	0.839	"AI language models help me improve the quality of learning English writing."
Students' PEOU	6	0.706	"I'm able to use AI language model platforms easily to achieve the requirements in my articles."
Benefits	10	0.915	"My writing skills improved after using AI language models."
Drawbacks	10	0.857	"AI language models make me lazy to think of new ideas because they come up with a lot more creative ideas."

Cronbach's Alpha is used to analyze the survey's reliability. The rule of an acceptable range of Cronbach's alpha is a value above 0.70, 0.80 or higher is preferred (Cortina, 1993). At first, a pilot study among 40 students was conducted, using the original version of the questionnaire with 35 items. After testing the pilot data on SPSS, two items were eliminated from the section about students' PEOU, reducing the number of items of this section from 8 to 6. This is because those two items made the Cronbach's Alpha of the set be at 0.697, which was under the recommended Cronbach's Alpha 7.0. The revised questionnaire then featured better Cronbach's Alpha as shown in the table above. All the items have Cronbach's alpha above 0.70, which shows that reliability is at an acceptable level.

To assess students' perceptions of using AI language models, the researcher designed a questionnaire with a 5-point Likert scale and the items based on the TAM framework with two main factors PU and PEOU. The questionnaire is divided into four sections, each of which answers three research questions in this study. Sections 1 and 2 will answer research question 1 regarding students' views of use, section 3 will answer research question 2 about the specific benefits of AI language models, and section 4 will answer question 3 about the disadvantages of utilizing AI language models.

The study's validity was confirmed through evaluations by instructors and survey participants. A lecturer with research experience reviewed the study's content, procedures, and data analysis to verify the accuracy and reliability of the results, which made significant contributions to

confirming the research methodologies and outcomes. Additionally, all responses from survey participants were also used to check the accuracy and reliability of the data. This includes reviewing the responses, evaluating the dependability of the information gathered, and finding out about any problems that happened during the data collection process.

To complement the quantitative data and gain a deeper understanding of participant perspectives, semi-structured interviews were used. This approach utilizes a core set of pre-determined questions, in this case, six questions in total, that serve as a guide, while allowing flexibility for follow-up inquiries and exploration of emerging topics (Gill et al., 2008). This flexibility fosters a more natural conversation, encouraging participants to share their opinions, feelings, beliefs, and motivations regarding the research topic (Gill et al., 2008).

Data Collection

Distributing Questionnaires

Google Forms was used in this study's survey questionnaire. Using an online questionnaire created by the Google Forms tool is the best option for both participants and researchers because it makes the process more convenient (Abhishek, 2024; Vasantha & Harinarayana, 2016). The questionnaire had four main sections based on three research questions. Parts 1 and 2 include the observations about students' attitudes toward AI language models' benefits on writing skills. Part 3 explores the benefits of AI language models for students' writing learning. Part 4 focuses on exploiting the disadvantages when students use AI tools in the writing classroom. All format sections were a 5-point Likert scale (strongly disagree, disagree, no opinion, agree, strongly agree) to elicit participants' opinions. Once the researcher receives the required responses from participants, the Google Form will close for data collection and analysis.

Conducting Interviews

There are a total of six interview questions, with follow-up ones. With the support of experienced researchers, questions are checked, reviewed, confirmed and adjusted to ensure their relevance to the research. The clarification of the questions is also considered to avoid complications during the interview process and later data analysis.

Interviews are conducted and recorded using the Microsoft Teams application. This app allows organizing a meeting, recording it, and also creating a script conversation instead of meeting in person to conduct interviews and take notes (Sah et al., 2020; Wakelin et al., 2024). All of these operations can be performed on the application.

The researcher asked for the participants' agreement before recording the interview and committed all the information for the aim of the study. The responses are transcribed using Teams' transcript tool. The responses are returned to the respondents for confirmation of what they said throughout the interviews. In the interview process, respondents can refuse to answer questions or can ask to stop at any time during the interview.

Data Analysis

Quantitative data were processed through SPSS software, which was described and analyzed by using basic statistics such as Mean (M) and Standard Deviation (SD) in the Likert scales. The study of Sin Yin et al. (2016) revealed that in Likert scales, M and SD ensure grouped items provide similar information. As suggested by Alkharusi (2022), the interpretation of M would be:

- Mean: 1.00 – 1.49: Strongly disagree
- Mean: 1.50 – 2.49: Disagree

- Mean: 2.50 – 3.49: Neutral
- Mean: 3.50 – 4.49: Agree
- Mean: 4.50 – 5.00: Strongly agree

Meanwhile, qualitative data from semi-structured interviews was analyzed to provide deeper insights and complement the findings from the quantitative data, using methods outlined by Creswell (2014). A thematic analysis approach was employed to identify recurring patterns, themes, and key insights from participants' responses. The analysis followed a systematic process, including data familiarization, coding, theme development, and interpretation, ensuring a rigorous and structured examination of the qualitative data. By integrating both quantitative and qualitative findings, the study was able to offer a more comprehensive understanding of the research issues, allowing for a richer, more nuanced interpretation that clarifies and supplements the survey results.

Findings and Discussion

Table 2

AI Language Models That Students Have Used

AI language models	Total	Percentage
Google Translate	138	93.9
Chat-GPT 3.5	99	67.3
Grammarly	128	87.1
Quillbot	108	73.5
Gemini	49	33.3
Bing Translator	1	0.7
Papago	1	0.7
TFlat	1	0.7

Table 2 presents the AI language models that students have used. The most frequently used models were Google Translate (93.9%), Grammarly (87.1%), Quillbot (73.5%), and ChatGPT 3.5 (67.3%). Gemini was used by 33.3% of the respondents, while Bing Translator, Papago, and TFlat were each used by 0.7%.

Table 3

Students' Attitudes of Using AI Language Models

		n	M	S.D.
1	I find it difficult to learn writing without using AI language models.	147	3.10	0.96
2	The process of writing English learning is easy when I use AI language models.	147	3.98	1.10
3	AI language models help me improve the quality of my essays.	147	4.14	0.77
4	AI language models help increase the quality of writing English learning.	147	3.88	0.87
5	Using AI language models helps me save time and complete exercises more quickly.	147	4.23	0.82
6	AI language models help me increase my proactiveness in writing learning.	147	3.44	1.11
7	Overall, I believe AI language models are useful for my learning writing process.	147	4.19	0.78

The results in Table 3 showed that most of the respondents do not need to use AI language models in all their writing exercises. They can finish writing tasks by themselves. Also, they already know how they can benefit from AI tools for checking their writing and saving time, and they believe that these applications are useful for them (M=4.19, SD=0.78). As revealed in the interview, students #1, #4, #7, #8, and #10 said that they would use AI tools, such as Grammarly, ChatGPT, and Google Translate, to check grammar, vocabulary and the fluency of their writing after finishing their work. However, the opinion of respondents who do not believe that AI tools can increase their proactiveness is varied (M=3.44, SD=1.11). For example, students #2 and #3 have mentioned that they intend to be lazy for thinking new ideas for their writing.

> *"I need to learn how to use these applications properly because I recognize that I immediately check everything I have written" (student #2).*

Table 4

Students' Perceptions of Using AI Language Models

		n	M	S.D.
1	I'm confused when using AI language models.	147	2.50	1.05
2	The AI language model systems are still stereotypical and lack flexibility.	147	3.33	1.10
3	I need to check the instructions for using AI language models regularly.	147	3.10	1.10
4	The features of AI language models are difficult to use.	147	2.50	1.01
5	I can operate on AI language model platforms to easily serve my articles' requirements.	147	4.01	0.81
6	AI language model platforms have instructions for users to use the features.	147	3.61	0.99

Table 4 shows that most students find these AI language models easy to use, and they do not need to spend time searching for the instructions. All the students in the interview agree that these tools have user-friendly interfaces, and they can access them easily. Students #1, #3, #8, and #9 mentioned that they only need to register with their email accounts and then use these tools; some apps do not require registration.

However, students #1, #2, #4, #7, #8, and #10 said they would access some platforms to find the proper prompts to use ChatGPT. They mentioned that sometimes ChatGPT could not reply to their answer accurately because they did not clarify their requirements. According to students #2 and #10, the prompts need to be checked to achieve the answers that serve their requirements because these tools often offer us general answers and cannot solve their problems.

As shown in Table 5, all the students agree that AI language models are beneficial virtual assistants as the M of all items is above 3.49 and the SD of all items is below 1. The majority of students perceive they can benefit from these AI tools. Their writing style, grammar, and vocabulary are improved based on the AI feedback.

Table 5

The Benefits of AI Language Models

		n	M	S.D.
1	AI has a large amount of information that helps me in the writing process.	147	4.21	0.86
2	AI language models help me learn new vocabulary.	147	3.99	0.94
3	I can check my grammar thanks to AI language models.	147	4.18	0.79
4	AI language models help me arrange sentences in an essay logically.	147	3.90	0.95
5	AI language models give me ideas for my writing topic.	147	4.05	0.82
6	I paraphrase quickly by using AI language models.	147	4.23	0.79
7	AI language models help me adjust the appropriate writing style.	147	4.03	0.80
8	I complete my writing assignments quickly and easily by using AI language models.	147	4.10	0.86
9	The reviews and comments by AI language models are helpful for my writing.	147	3.98	0.88
10	My writing skills improved after using AI language models.	147	3.90	0.88

However, according to students #2 and #4, they gradually acquired the AI comments without checking but sometimes the feedback was unnecessary. Also, most of the interview respondents have used AI tools for summarizing and said that the summary versions have not met their requirements.

> *"I have used these AI tools to brainstorm ideas for my essays; however, the ideas are general. I also have to use more than two tools for paraphrasing and searching information to get accurate outputs" (student #4).*

Table 6

The Drawbacks of AI Language Models

		n	M	S.D.
1	I depend on using AI language models to check grammar.	147	3.37	1.07
2	I depend on using AI language models to check vocabulary.	147	3.28	1.06
3	I had to adjust the prompts many times to achieve the exact answer I asked from AI language models.	147	3.75	0.97
4	I depend on using AI language models to paraphrase my essay.	147	3.50	1.01
5	AI language models make me lazy to think of new ideas because they come up with more creative ideas.	147	3.49	1.11
6	Relying on AI language models exposes me to the risk of unintentionally plagiarizing content.	147	3.65	1.04
7	I'm worried that my personal information and data will be leaked and used by AI language models.	147	3.60	1.12
8	The answers of AI language models are not completely accurate.	147	3.92	0.95
9	Comments from AI language models are still general and do not contribute much to improving my writing.	147	3.64	1.06
10	I'm worried that I will cheat on most writing assignments because of the AI convenience.	147	3.50	1.12

As revealed in Table 6, the responses show that students have not recognized whether they depend on AI tools to check grammar and vocabulary, but these opinions are varied, which should be focused for further discussion (M=3.37, SD=1.07; M=3.28, SD=1.06). Also, most students have no opinion about AI making them lazier (M=3.49, SD=1.11).

> *"The ideas that come from AI outputs are not human-like answers and do not serve my requirements, I need to check the information before using them in my essay" (student #4).*

Most of the respondents show that they choose to build their writing ideas by themselves without using AI due to their lack of creativity. These tools play the role of assistants in providing suggestions for grammar and word choice. The students most concerned about utilizing AI tools can lead to unintentional plagiarizing and creating dependence. The comments from the chatbots have not been clarified, so they do not contribute much to students' writing.

Furthermore, participants were not aware that personal information had been leaked. According to most students, these chatbots only access their email accounts, and they can use the features immediately. Nonetheless, student #2 mentioned that all the activities people have done on the Internet will leave footprints, so they should pay more attention to their data. Student #8 also recommends that people should check the information carefully whenever using the AI tools because the prompts might include their personal data.

Discussion

The finding that students frequently utilize Google Translate, Grammarly, Quillbot, and ChatGPT 3.5 aligns with the literature highlighting the widespread integration of large AI language models in language learning (Kasneci et al., 2023; Alharbi, 2023; Jumriah et al., 2024). The positive attitudes reported by students regarding the usefulness of AI in saving time and improving essay quality resonate with the benefits of AI as writing assistants discussed by Athanassopoulos et al. (2023) and the advantages of chatbots in providing readily available assistance noted by Sumakul et al. (2022).

Beyond accessibility, the positive attitudes reported by students regarding the usefulness of AI in saving time and improving essay quality align with the benefits of AI as writing assistants discussed by Athanassopoulos et al. (2023) and the advantages of chatbots in providing readily available assistance noted by Sumakul et al. (2022). The preference observed in this study for students to generate their own ideas and use AI for grammar and word choice further suggests a recognition of AI as a tool for refinement, consistent with Godwin's (2022) perspective on the challenges of providing truly beneficial corrective feedback in second-language writing, where AI can offer immediate suggestions on certain aspects.

Despite these advantages, several challenges and concerns emerged from this study. Firstly, the concern raised by some students in this study about the potential for AI to induce laziness in generating ideas is similar to the ethical concerns and potential drawbacks discussed by Tira (2023), who mentioned the threat of technology leading to easier cheating and plagiarism, and the reduction of cognitive effort in language processing. Secondly, the challenges in formulating effective prompts for ChatGPT and the issue of inaccurate or general responses demonstrates the "hallucination" phenomenon reported in LLMs by Athaluri (2023), Lee et al. (2022), Tira (2023), and Ziwei et al. (2023), and the problem of inaccurate responses noted by Phan (2023). Furthermore, the instances of students in this study accepting AI feedback without critical evaluation align with the concern raised by Duong & Le (2024) and Ha & Ho (2025) about tools like Grammarly potentially misleading students into focusing solely on grammatical

correctness and overlooking crucial aspects of writing like logic and style. Finally, the concerns about potential dependence on AI and the risk of unintentional plagiarism expressed by students in this study mirror the issues highlighted by Kasneci et al. (2023) and Xiao and Zhi (2023) regarding the ease with which AI-generated text can be submitted as one's own, raising ethical considerations in academic settings. Finally, the limited awareness regarding personal data privacy among some students in this study contrasts with the ethical and professional apprehensions discussed in the literature, where concerns about data security and privacy breaches are emerging as critical issues (Burkhard, 2022).

Conclusion

Summary of the Study

From the information analyzed, students' perceptions of AI language models are consistent with what is discussed in the literature review. Most students find AI language models easy to interact with, and they do not encounter any significant problems accessing these tools.

Using AI tools offers them substantial benefits in the process of learning writing. Firstly, these virtual assistants help students save time by providing a vast amount of information quickly. Additionally, students agree that using AI language models improves their vocabulary, grammar, and writing style. The diverse vocabulary and writing ideas generated by AI help students expand and enrich their vocabulary. Moreover, these virtual assistants provide students with instant feedback, help them identify the mistakes and shortcomings in their work promptly, and enhances their writing abilities in the long term. Last but not least, AI tools increase students' confidence and accuracy in presenting their opinions.

However, the study also highlights a potential downside: students' dependence on AI technology. Many students rely on AI to check their writing before submitting it to their teachers. This raises questions about the development of students' autonomy. While using AI, students should be mindful of not becoming too dependent on technology and should focus on developing their self-learning skills.

Overall, AI language models have brought students numerous benefits in learning to write. They provide diverse information about vocabulary and writing ideas, help improve writing, and support grammar checking. However, students need to create accurate prompts to get the best results from AI. They should also be cautious not to rely excessively on technology and should work on improving their writing skills through self-study and practice.

Limitations of the Study

This study has several limitations, including a small sample size and reliance on manual analysis for qualitative data, which may not fully eliminate deviations and may affect the level of trust in the findings. Future studies should aim for larger sample sizes, use random sampling techniques, and consider alternative research methods beyond descriptive studies to generate deeper and more comprehensive insights in the field.

Implications

The study offers several implications for students and instructors. Students must learn how to create effective prompts to achieve accurate results from AI models. Offering detailed requirements is crucial for obtaining favorable responses from AI. Despite the diversity of information provided by AI, concerns about accuracy and detail remain. Therefore, students need the knowledge and analytical skills to use AI-generated information reasonably and

reliably.

For teachers, AI language models hold potential benefits due to their convenience and time-saving capabilities. Instructors should explain to students the role and capabilities of AI in the learning process, helping students use these models effectively. AI tools can assist teachers in designing lessons, updating information quickly, and recommending classroom activities. Appropriately applying AI tools can foster positive interactions with students. Since AI chatbot outputs may not always be accurate, instructors should encourage students to analyze, evaluate, and select information critically, thereby developing their critical thinking skills.

This study underscores the need for further research. Future studies should explore the long-term impacts of using AI language models in writing instruction and investigate strategies to mitigate students' dependence on technology. Additionally, research should focus on how AI tools can be integrated into different educational contexts and disciplines, examining their effects on various aspects of learning and teaching. Experimental designs with larger and more diverse samples can provide more robust evidence and deeper insights, helping to refine the use of AI in educational settings.

Acknowledgments

We sincerely extend our heartfelt appreciation to Assoc. Prof. Pham Vu Phi Ho and Mr. Le Quang Thao for their invaluable suggestions and constructive feedback, which significantly contributed to the study's design. Additionally, we express our deepest gratitude to the students at Van Lang University, particularly those from the Faculty of Foreign Languages, for their active participation in the questionnaire and their willingness to engage in the interview process. Their insights and support were instrumental in shaping the findings of this research.

References

Abhishek, J. 2024. Google Forms. In P. Ashish, K.M. Amit & J. Abhishek (Eds.), *Open electronic data capture tools for medical and biomedical research and medical allied professionals* (pp. 331-378). Academic Press. https://doi.org/10.1016/B978-0-443-15665-6.00008-7.

Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it? *Indian Journal of Medical Specialities, 4*(2). https://doi.org/10.7713/ijms.2013.0032

ALAfnan, M. A., Samira Dishari, Marina Jovic, & Koba Lomidze. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. *Journal of Artificial Intelligence and Technology, 3*(2), 60–68. https://doi.org/10.37965/jait.2023.0184

Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, 2023, 1–15. https://doi.org/10.1155/2023/4253331

Alkharusi, H. (2022). A descriptive analysis and interpretation of data from Likert scales in educational and psychological research. . *Indian Journal of Psychology and Education*, *12*(2), 13-16

Almalki, S. (2016). Integrating quantitative and qualitative data in mixed methods research—Challenges and benefits. *Journal of Education and Learning, 5*(3), 288.

https://doi.org/10.5539/jel.v5n3p288

Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, *15*(4), e37432. https://doi.org/10.7759/cureus.37432

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research, 3*(2), 818–824. https://doi.org/10.25082/amler.2023.02.009

Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate.* arXiv.org. https://arxiv.org/abs/1409.0473v7

Baskara, F. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning, 5*(1), 44-55. doi:https://doi.org/10.31763/ijele.v5i1.858

Bender, E M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *The 2021 ACM Conference on Fairness, Accountability, and Transparency.* https://doi.org/10.1145/3442188.3445922

Bhandari, P. (2020, June 19). *What is qualitative research?* | Methods & Examples. Scribbr. https://www.scribbr.com/methodology/qualitative-research/

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English With Technology, 2023*(1). https://doi.org/10.56297/bkam1691/wieo1749

Burkhard, M. (2022, November 22). *Student perceptions of AI-powered writing tools: Towards individualized teaching strategies* [Conference presentation abstract] .19th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2022), https://eric.ed.gov/?id=ED626893

Carvalho, L., Martinez-Maldonado, R., Tsai, Y. S., Markauskaite, L., & De Laat, M. (2022). How can we design for learning in an AI world? *Computers and Education: Artificial Intelligence, 3*, 100053. https://doi.org/10.1016/j.caeai.2022.100053

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science, 136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

Chu, Q. B., Nguyen, M. B., & Nguyen, A. P. (2024). Using Artificial Intelligence Tool in Studying English Skills in Vietnam – An Experimental Research for Vietnamese High School Students. *Journal of Ecohumanism*, *3*(6), 1883–1894. https://doi.org/10.62754/joe.v3i6.4144

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037//0021-9010.78.1.98

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches.* SAGE Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Duong, T. K. H., & Le , T. T. H. (2024). Utilizing Artificial Intelligence in Writing Feedback: Benefits and Challenges for First-Year Students at Hanoi University of Industry. *Proceedings of the AsiaCALL International Conference*, *6*, 238–249. https://doi.org/10.54855/paic.24617

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching, 12*(1), 44–58. https://doi.org/10.15294/elt.v12i1.64069

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(15)*,* 1-15. https://doi.org/10.2139/ssrn.4358789

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. 2011. *How to design and evaluate research in education.* McGraw-Hill.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008, March). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal, 204*(6), 291–295. https://doi.org/10.1038/bdj.2008.192

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology, 26*(2), 5–24. http://doi.org/10125/73474

Golzar, J., Noor, S., & Tajik, O. (2022). Convenience Sampling. *International Journal of Education &Language Studies*, *1*(2), 72-77. https://doi.org/10.22034/ijels.2022.162981

Ha, Y. N., & Ho, N. P. (2025). EFL postgraduate students' perceptions on the use of Grammarly and peer feedback to improve their academic writing skills. *International Journal of TESOL & Education, 5*(1), 25-49. https://doi.org/10.54855/ijte.25512

Ho, P. X. P. (2024). Using ChatGPT in English language learning: A study on I.T. students' attitudes, habits, and perceptions. *International Journal of TESOL & Education, 4*(1), 55-68. DOI: https://doi.org/10.54855/ijte.24414

Ji, Z., Yu, T., Xu, Y., Lee, N., Ishii, E., & Fung, P. (2023). Towards mitigating LLM hallucination via self reflection. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 1827–1843). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.123

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396-403. https://doi.org/10.9734/BJAST/2015/14975.

Jumriah, J., Susilawati, E. S., Supriatna, E., Smas, M. H., & Arini, I. (2024). Analysis of the use of ChatGPT to improve student performance. EDUKASIA: *Jurnal Pendidikan Dan Pembelajaran, 5*(1), 127-136. https://doi.org/10.62775/edukasia.v5i1.725

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023, April). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Katar, O., Özkan, D., -3, G., Yildirim, Z., & Acharya, U. R. (2023). Araştırma Makalesi Yazımında GPT-3 Yapay Zeka Dil Modeli Değerlendirmesi. *Turkish Journal of Science and Technology, 18*(2), 311–318. https://doi.org/10.55525/tjst.1272369

Kelle, U. (2006). Combining qualitative and quantitative methods in research practice: purposes and advantages. *Qualitative Research in Psychology*, *3*(4), 293–311. https://doi.org/10.1177/1478088706070839

Li, J. (2024). The evolution, applications, and future prospects of large language models: An in-depth overview. *Applied and Computational Engineering, 35*(1), 234–244. https://doi.org/10.54254/2755-2721/35/20230399

Linguaserve (2023, November 24). *What is neural machine translation (NMT) and how can it help you*? https://www.linkedin.com/pulse/what-neural-machine-translation-nmt-how-can-help-you-linguaserve-nulpf/

Marzuki, A., Widiati, U., Rusdin, D., Darwin, , & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2). https://doi.org/10.1080/2331186X.2023.2236469

Micheni, E., Machii, J., & Murumba, J. (2024). The role of Artificial Intelligence in education. *Open Journal for Information Technology, 7*(1), 43-54. https://doi.org/10.32591/coas.ojit.0701.04043m

Miller, J. & Khera, O. 2017. Digital library adoption and the technology acceptance model: A cross-country analysis. *The Electronic Journal of Information System in Developing Countries, 40*(1), 1-19. https://doi.org/10.1002/j.1681-4835.2010.tb00288.x

Nguyen, H. H. B., Ngoc, H. H. B., & Dan, T. C. (2024). EFL Students' Perceptions and Practices of Using ChatGPT for Developing English Argumentative Essay Writing Skills. *European Journal of Alternative Education Studies, 9*(1), 168-216. http://dx.doi.org/10.46827/ejae.v9i1.5341

Nguyen, T. T. H. (2023). EFL Teachers' Perspectives toward the Use of ChatGPT in Writing Classes: A Case Study at Van Lang University. *International Journal of Language Instruction, 2*(3), 1-47. https://doi.org/10.54855/ijli.23231

Novarita, N. (2021). The students' perception toward using mobile technology for learning English. *Edukasi Lingua Sastra, 19*(1), 74–82. https://doi.org/10.47637/elsa.v19i1.326

Pham, T. T., Nguyen , L. A. D., Dang, H. M., & Le, T. T. P. (2024). Exploring Tertiary Vietnamese EFL Students' Engagement in Vocabulary Learning through the Use of an AI Tool. *Proceedings of the AsiaCALL International Conference*, *4*, 129–149. https://doi.org/10.54855/paic.23410

Phan, T. N. L. (2023). Students' perceptions of the AI technology application in English writing classes. *AsiaCALL International Conference, 4*, 45–62. https://doi.org/10.54855/paic.2344

Roose, K. (2023, January 12) *Don't ban ChatGPT in schools. Teach with it*., The New York Times. https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html?smid=url-share

Sah, L. K., Singh, D. R., & Sah, R. K. (2020). Conducting qualitative interviews using virtual communication tools amid Covid-19 pandemic: A learning opportunity for future

research. *JNMA; Journal of the Nepal Medical Association*, *58*(232), 1103–1106. https://doi.org/10.31729/jnma.5738

Salloum, S. A., Alhamad, A. Q. M., Al-Emran, M., Monem, A. A., & Shaalan, K. (2019). Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model. *IEEE Access, 7*, 128445-128462. https://doi.org/10.1109/ACCESS.2019.2939467

Sari, A. N. (2024). Exploring the Potential of Using AI Language Models in Democratising Global Language Test Preparation. *International Journal of TESOL & Education, 4(*4), 111-126. https://doi.org/10.54855/ijte.24447

Sin Yin, T., Othman, A. R., Sulaiman, S., Mohamed-Ibrahim, M. I., & Razha-Rashid, M. (2016). Application of mean and standard deviation in questionnaire surveys: Construct validation. *Jurnal Teknologi*, *78*(6-4), 99-105. https://doi.org/10.11113/jt.v78.8983

Sumakul, Y. G., Hamied, F. A., & Sukyadi, D. (2022). Students' perceptions of the use of AI in a writing class. *67th TEFLIN International Virtual Conference & the 9th ICOELT 2021 (TEFLIN ICOELT 2021)*. https://doi.org/10.2991/assehr.k.220201.009

Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open, 1*, 5–21. https://doi.org/10.1016/j.aiopen.2020.11.001

Tran, T. T. H. (2024). AI Tools in Teaching and Learning English Academic Writing Skills. *Proceedings of the AsiaCALL International Conference, 4,* 170–187. https://doi.org/10.54855/paic.23413

Vasantha Raju, N. N. S. H., & Harinarayana, N. S. (2016, January). *Online survey tools: A case study of Google Forms* [Conference presentation]. National conference on Scientific, Computational & Information research trends in engineering, GSSS-IETW, Mysore, India.

Vasantha Raju, N., Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Wakelin, K. J., McAra-Couper, J., & Fleming, T. (2024). Using an Online Platform for Conducting Face-To-Face Interviews. *International Journal of Qualitative Methods*, *23*, 1-9. https://doi.org/10.1177/16094069241234183

Wittink, D.R. & Bayer, L.R.. (2003). *The Measurement Imperative, 6*. 14-23. https://www.researchgate.net/publication/292779687_The_Measurement_Imperative

Wu, Y., Schuster, M., Chen, Z., Le, Q., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kudo, T., Kazawa, H., & Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation.* arXiv.org. https://doi.org/10.48550/arXiv.1609.08144

Xiao, Y., & Zhi, Y. (2023). An Exploratory Study of EFL Learners' Use of ChatGPT for language learning tasks: Experience and perceptions. *Languages, 8*(3), 212. https://doi.org/10.3390/languages8030212

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., & Hu, X. (2023, April 26).

Harnessing the power of LLMs in practice: a survey on ChatGPT and beyond. arXiv.org. https://arxiv.org/abs/2304.13712v2

Ziwei, J., Nayeon, L., Rita, F., Tiehzheng, Y., Dan, S., …Pascal, F. (2022). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1-38. https://doi.org/10.1145/3571730

Biodata

Pham Ngoc Thai Binh, a former English Language major at Van Lang University, possesses three years of experience as a teaching assistant in an English center. Her professional interests include a passion for languages, a dedication to teaching, and a keen interest in blended learning methodologies.

With experience as an English Language Teaching lecturer at Van Lang University, Ms. Tran Thi Mai has taught a diverse range of courses, from general English to specialized subjects and test preparation. Her teaching is driven by a philosophy of constant enhancement, aiming to equip students with enriched knowledge, honed skills, and a positive, independent attitude towards lifelong learning. Her research explores mindful learning, blended learning, and autonomy in language acquisition.

APPENDICES

APPENDIX 1: QUESTIONNAIRE

Dear Participants,

My name is Phạm Ngọc Thái Bình, and I am a final-year student (Cohort 26) majoring in English Language at the Faculty of Foreign Languages, Van Lang University. I am currently conducting a research study titled:

"Students' Perceptions of AI Language Models as Virtual Assistants in Learning Writing."

What Are AI Language Models?

AI language models are artificial intelligence (AI) systems trained to process and generate human-like text. These models, such as ChatGPT, Gemini (Bard), Google Translate, and Grammarly, learn from large and diverse datasets to generate new text based on their training. They have various applications, including content creation, automated email responses, coding assistance, storytelling, translation support, and other natural language processing tasks.

Survey Participation & Confidentiality

- All personal information provided in this survey will remain strictly confidential.
- The survey is conducted for academic purposes only and is non-commercial.
- As this research contributes to my graduation thesis, I kindly ask for 10 to 20 minutes of your time to carefully read and answer the questions.
- The success of this study relies on your valuable responses and participation, which will greatly support the research findings.

Thank you for taking the time to participate in this survey! I truly appreciate your support and hope you achieve all your goals in the future.

For any inquiries, please feel free to contact me at: [contact detail].

Part 1. Personal information

1. What's your major at Van Lang University? _____

2. You are a: ☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior

3. Have you ever used AI language models in learning languages? ☐ Yes ☐ No

4. Which AI language models do you often use to assist in language learning? (can choose multiple answers):

☐ Google Translate ☐ Chat-GPT 3.5 ☐ Grammarly ☐ Quillbot ☐ Gemini

☐ Others: ….

Please read each one and click on the bullet to show how much you agree or disagree with each statement.

1: Strongly disagree
2: Disagree
3: Neutral
4: Agree
5: Strongly agree

Part 2. Students' attitudes of using AI language models

		1	**2**	**3**	**4**	**5**
1	I find it is difficult to learn writing without using AI language models.					
2	The process of writing English learning is easy when I use AI language models.					
3	AI language models help me improve the quality of my essays.					

4	AI language models help increase the quality of learning English writing.					
5	Using AI language models helps me save time and complete exercises more quickly.					
6	AI language models help me increase my proactiveness in writing learning.					
7	Overall, I believe AI language models are useful for my learning writing process.					

Part 3. Students' perceptions of using AI language models

		1	**2**	**3**	**4**	**5**
1	I'm confused when using AI language models.					
2	The AI language model systems are still stereotypical and lack flexibility.					
3	I need to check the instructions for using AI language models regularly.					
4	The features of AI language models are difficult to use.					
5	I can operate on AI language model platforms to easily serve my articles' requirements.					
6	AI language model platforms have instructions for users to use the features.					

Part 4. The benefits of AI language models

		1	**2**	**3**	**4**	**5**
1	AI has a large amount of information that helps me in the writing process.					
2	AI language models help me learn new vocabulary.					
3	I can check my grammar thanks to AI language models.					
4	AI language models help me arrange sentences in an essay logically.					
5	AI language models give me ideas for my writing topic.					
6	I paraphrase quickly by using AI language models.					
7	AI language models help me adjust the appropriate writing style.					
8	I complete my writing assignments quickly and easily by using AI language models.					
9	The reviews and comments by AI language models are helpful for my writing.					
10	My writing skills improved after using AI language models.					

Part 5. The drawbacks of AI language models

		1	2	3	4	5
1	I depend on using AI language models to check grammar.					
2	I depend on using AI language models to check vocabulary.					
3	I had to adjust the prompts many times to achieve the exact answer I asked from AI language models.					
4	I depend on using AI language models to paraphrase my essay.					
5	AI language models make me lazy to think of new ideas because they come up with more creative ideas.					
6	Relying on AI language models exposes me to the risk of unintentionally plagiarizing content.					
7	I'm worried that my personal information and data will be leaked and used by AI language models.					
8	The answers of AI language models are not completely accurate.					
9	Comments from AI language models are still general and do not contribute much to improving my writing.					
10	I’m worried that I will cheat on most writing assignments because of the AI convenience.					

APPENDIX 2: RELIABILITY TEST

1. Students' attitudes of using AI language models

Case Processing Summary

		N	%
Cases	Valid	111	100.0
	Excluded[a]	0	.0
	Total	111	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.839	7

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
A1	23.8378	15.864	.348	.855
A2	22.9189	14.184	.748	.794
A3	22.7928	14.548	.664	.806
A4	23.0000	14.255	.673	.804
A5	22.7027	14.847	.617	.813
A6	23.4414	13.885	.506	.838
A7	22.7658	14.636	.697	.803

2. Students' perceptions of using AI language models

2.1. Original items' reliability

Case Processing Summary

		N	%
Cases	Valid	111	100.0
	Excluded[a]	0	.0
	Total	111	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.697	8

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
B1	24.6486	14.975	.376	.670
B2	23.8649	13.700	.505	.637
B3	24.0270	14.117	.448	.652
B4	24.5766	14.028	.495	.641
B5	23.1532	16.331	.318	.682
B6	23.5225	15.088	.377	.670
B7	23.2342	16.308	.280	.689
B8	23.0450	16.662	.274	.689

2.2. Revised items' reliability:

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	111	100.0
	Excluded[a]	0	.0
	Total	111	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.706	6

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
B1	16.6216	10.274	.536	.635
B2	15.8378	9.773	.563	.623
B3	16.0000	9.673	.583	.616
B4	16.5495	9.741	.613	.608
B5	15.1261	13.638	.094	.749
B6	15.4955	12.198	.235	.727

3. The benefits of AI language models

		N	%
Cases	Valid	111	100.0
	Excluded[a]	0	.0
	Total	111	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.915	10

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
C1	36.2883	34.280	.661	.908
C2	36.4865	32.888	.697	.906
C3	36.2883	34.716	.655	.908
C4	36.6216	33.201	.671	.907
C5	36.4775	36.724	.443	.919
C6	36.2252	34.031	.735	.904
C7	36.4234	34.028	.743	.903
C8	36.4234	32.792	.780	.900
C9	36.4865	32.979	.785	.900
C10	36.5766	34.192	.699	.905

4. The drawbacks of AI language models

Case Processing Summary

		N	%
Cases	Valid	111	100.0
	Excluded[a]	0	.0
	Total	111	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.857	10

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
C1	32.1441	41.179	.585	.842
C2	32.2973	41.774	.545	.846
C3	31.8018	42.833	.512	.848
C4	32.0000	41.491	.605	.841
C5	32.0991	40.472	.598	.841
C6	31.9009	41.345	.579	.843
C7	32.0000	40.491	.569	.844
C8	31.5856	41.809	.599	.841
C9	31.8829	41.850	.552	.845
C10	32.0721	41.886	.500	.850

Language and Knowledge Content Gaps in Students' Academic Writing: A Case Study in University of Science and Technology- The University of Da Nang

Nguyen Thi Tu Trinh[1*]

[1]University of Science and Technology – The University of Da Nang, Da Nang City, Vietnam
[*]Corresponding author's email: ntttrinh@dut.udn.vn
[*]https://orcid.org/0000-0002-0353-9672
https://doi.org/10.54855/ijte.25526

Received: 25/01/2025 Revision: 05/05/2025 Accepted: 12/05/2025 Online: 13/05/2025

ABSTRACT

This study aims to identify and bridge the major language and knowledge gaps existing in their IELTS writing task 2. A collection of 202 essays made by engineering freshmen at the University of Technology- The University of Da Nang (DUT-UDN) is interpreted and organized into two strata: (i) lexico-grammar and (ii) knowledge content stratum. The descriptive and qualitative analysis of the collected data reveals that there is a tremendous number of grammatical errors, such as tenses, conjunctions, prepositions, and others, as well as wrong word use or lexical errors on account of word-by-word translation. Particularly, this study sheds light on the challenges of a persistent absence of experience and real-world knowledge in supporting techniques to offer solid persuasion to readers within a cross-cultural perspective. It is concluded that besides teaching grammatical and lexical ranges, introducing, practicing, and reinforcing argumentation, critical reasoning, typical exemplifying, quoting, and citing statistical evidence techniques are also essential in academic essays.

Keywords: Language and Knowledge Content Gaps, Academic Writing, Cross-Cultural Perspective

Introduction

The English language testing is faced with surging demands of accountability in all language examinations offered to the Vietnamese public. It is expected to give well-grounded measures to test takers, and the International English Language Testing System (IELTS) appears to meet the public's expectations. Over 30 years, IELTS has developed and undergone numerous changes, as well as been documented by Taylor and Falvey (2007), Davies (2008), and Weir and O'Sullivan (2017). The primary test of IELTS was the English Language Testing Service (ELTS), first introduced in 1980 by the British Council, in collaboration with the University of Cambridge Local Examinations Syndicate (UCLES) (now Cambridge Assessment English). ELTS was developed as an instrument for the Council to assess the English proficiency of international students awarded by the British Government for study or training in the United

CITATION | Nguyen, T. T. T. (2025). Language and Knowledge Content Gaps in Students' Academic Writing: A Case Study in University of Science and Technology- The University of Da Nang. *International Journal of TESOL & Education*, *5*(2), 116-131. DOI: https://doi.org/10.54855/ijte.25526

Kingdom. Since its debut in 1989, IELTS, with a four-part test (Listening, Reading, Writing, and Speaking), has gained widespread recognition and is used in most higher educational institutions. There have been some changes to the IELTS writing task, but the basic structure has remained with one exception (Read, 2015). After long-standing debate and extensive research, the three discipline-based modules: A. Physical Science and Technology, B: Life and Medical Sciences, and C: Arts and Social Sciences for reading and writing were reduced to just a single Academic Module for students in all subject areas.

Teaching, learning, and grading writing skills requires mastery of punctuation, spelling, grammatical and lexical ranges and structure, stylistic and judgement skills. Students' ability to organize and develop ideas is a significant skill for real-life communication and is viewed as a backwash effect of teaching and learning. In other words, it requires a complete range of linguistic skills as well as a knowledge base. In this paper, short essays written by engineering freshmen are collected and interpreted from four angles: Grammar, Vocabulary, Coherence and Cohesion, and Task achievement. There has been an increasing interest in the study of challenges in English composition in general. However, fewer attempts have been made to comprehensively analyze language, knowledge, and real-world gaps. Particularly, the absence of knowledge and real-world gap in their IELTS writing task 2 is striking. Hopefully, this article will contribute to the study of L2 composition within linguistics and the cross-cultural field.

Literature Review

Language is commonly examined in terms of whether (a) spoken or written language and (b) productive and receptive skills. According to Badger (2024), writing and speaking are grouped together as productive skills, but the former is used in written languageand spoken language. Over the forty years of L2 composition study, much attention has been paid to interpreting challenges faced by L2 student writers that are attributed to their limited proficiency in the target language (Hamp-Lyons, 1991). Evan (2019) investigates the writing experiences and perceptions of Master's students, reveals their writing challenges, and makes recommendations on genre-based approaches, genre-sensitive pedagogy, and curriculum developers.

Nguyen and Nguyen (2022) conducted research on the problems and causes of taking IELTS writing task 2 for 205 IELTS learners from two centers in Ho Chi Minh City, Vietnam. The study outlines 12 problems and causes encountered by the participants and highlights the five most frequent problems, namely time constraint, paraphrasing, insufficient background knowledge, linking ideas, and writing confusing sentences, and four causes, such as Vietnamese influence, writing anxiety, lack of English exposure, and cultural barriers. Nghi and Truong (2023) explore the difficulties Vietnamese students encounter when learning academic writing in English and show a strong interest of most respondents in language learning, but their dissatisfaction and discontent with their English writing skills on account of insufficient practice time. Tran and Truong (2021) conducted a study on the frequencies of single-clause sentences and multi-clause sentences in IELTS writing task 2 and presented a close link between sentence types and IELTS composition band score. They point out that the lower the band score, the more single-clause sentences are available in a composition. The above studies focus

on analyzing grammatical range, accuracy, coherence, and cohesion. Moreover, fewer attempts have been made to find out and analyze experience and real-world knowledge gaps in IELTS writing task 2. Therefore, this study is fairly distinct from the previous works.

To fulfill the purpose of the study, the survey sought to answer the following research questions:

1. What are the language gaps in IELTS writing task 2 of engineering freshmen at DUT-UDN?
2. What are the experience and real-world knowledge gaps in IELTS writing task 2 for engineering freshmen in DUT-UDN?

Methods

Pedagogical Setting & Participants

With a population of 101 engineering students taking a two-credit English composition course at the University of Da Nang, all participants are volunteers, and there is no compensation for their participation. 89 male and 12 female freshmen aged 18 enrolling in two programs: Embedded System and Electrical Communication Engineering in DUT-UDN must fulfill the language requirement to achieve their engineering degree. To be specific, the minimum requirement is a score of IELTS band 6.0. All participants did the university placement tests and were granted B1 or B2 within the Vietnamese Six-level Foreign Language Proficiency Framework in 2020. Then they were invited to a diagnostic writing test and a final term writing test, with a response rate of 100 %. In addition, participants were requested to complete open questionnaires in Google Forms with a 100% response.

Design of the Study

Two hundred two short essays written by 101 engineering freshmen in the diagnostic tests and final-term tests at DUT-UDN are collected and analyzed in terms of four criteria suggested by the British Council, namely: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. They include 101 diagnostic writing tasks and 101 final-term writing tests. Vocabulary size plays a pivotal role in IELTS writing scores. Particularly, it outlines the language gap, experience, and real-world knowledge gap. In addition, 10 open-ended questionnaires in Google Forms are designed to determine their writing learning habits, strategies, attitudes, opinions, and beliefs about teaching and learning academic writing.

Data collection & analysis

This work adopts a descriptive and qualitative approach, and the data are analyzed and interpreted in terms of language gaps, experience gaps, and real-world knowledge gaps. The detailed discussion is outlined in Table 1.

Table 1

A detailed discussion of the study

Language gaps	Experience and real-world knowledge gaps
Lexical Resources	Typical examples
Grammar Range and Accuracy	Statistics
Coherence – Cohesion	Quotations

Findings

Research Question 1: What are the language gaps in IELTS writing task 2 of engineering freshmen in DUT-UDN?

Language gaps

Lexical Resources

Vocabulary Tiers

There is a school of thought that academic vocabulary is crucial in IELTS writing tests, and there should be some writing strategies to boost their academic vocabulary input. Milton et al. (2010) state, "The vocabulary size is the most important factor in determining success in writing, reading and listening and overall IELTS grades; more important than grammatical accuracy or the other factors" (p.96)

The three-tier vocabulary framework suggested by Beck et al. (2002) is exploited to analyze their writing in terms of general and academic vocabulary.

Tier 1 words

Tier 1 words are ubiquitous and frequently used in daily conversation and informal discourse. Therefore, these words are practical and functional but easy to understand and use. As a result, there is no need to teach or develop Tier 1 word vocabulary for them in IELTS writing.

Tier 2 words

With respect to academic vocabulary, Tier 2 words include high-utility words available in numerous academic domains and disciplines for intermediate English learners. It is vital to emphasize Tier 2 words in vocabulary instruction because of their valuable function and applicability to support their ideas, make arguments, and discuss evidence in their writing.

Tier 3 words

In distinction from Tier 2 words, Tier 3 words are more specialized and are notably absent in their IELTS writing part 2. In addition, Tier 3 words are low-frequency words limited to specific domains and fields (e.g., Chrysalis, filibuster, ambience, and others).

It is evident that academic English (Tier 2 and 3 words) deserves special attention because there is a strong correlation between students' vocabulary knowledge and their success as readers, learners, and even IELTS test takers. Nevertheless, the collected data has a highly limited range of academic vocabulary. The question is how to teach and assist them in building up and increasing the size of their academic English (Tier 2 and 3 words). There are three suggested

steps to guide them. Firstly, there is no doubt that IELTS reading passages can serve as an infinite source of academic vocabulary for them. Potential academic words are marked and selected to offer some instruction and information about them. Secondly, after mastering the words, let them rehearse and deepen their understanding of the new, varied uses and meanings by making new sentences to give them more chances to use these words in different collocations and contexts. Particularly, students are asked to construct their own general and academic vocabulary appendix during their prolonged learning process (See Table 2). It is considered a conscious process and action to recognize the distinction between these two categories and employ academic vocabulary in their writing.

Table 2

General and academic vocabulary appendix

General English (Tier 1 words)	Academic English (Tier 2 and 3 words)
A lot of/ Many	A plethora of Myriad A Multitude of
Always	Invariably
Available	Prevalent
Basic	Rudimentary
Benefits	Merits
Careful	Sobering
Change	Alter
Choose	Take a punt
Condition	Well-being
Crowded	Densely-populated
Difficult	Uphill
Enough	Sufficient
Exciting	Exhilarating
Extremist	Fanatical
Famous	Notorious
Friendly	Approachable
Frightening	Intimidating
Growth	Proliferation
Human	Homo sapiens
Lazy	Sedentary
Long	Prolonged
Near	In the proximity of
Perfect	Impeccable
Reason	Culprit
Rich	Well-off
Tiring	Tedious
Watchful	Vigilant

Schnoor (2003) claims that dictionaries are categorized into two kinds: monolingual versus bilingual dictionaries. The former gives an explanation of the word in one language, L1, while the latter provides direct translations from L1 to L2. To be specific, bilingual dictionaries involve L1 equivalents of L2 words. So, it is easy to look a word up in a bilingual dictionary.

Monolingual dictionaries, on the other hand, provide information only in the target language (Fan, 2000). The survey of dictionary preference shows that 82 % of participants strongly favor bilingual dictionary platforms such as Google Translate, Tflat, and others rather than monolingual counterparts or thesauruses that are less popular with learners on account of their complicated explanations. Most of the participants stated that finding words in bilingual dictionaries is much easier and quicker, but it does not give collocations. They also argue that bilingual dictionaries give more than one meaning of a word without using them in a context, as an example, which is confusing. For example, they have a vague idea about "polite" in the English–Vietnamese dictionary. "Polite" is translated as "lễ phép, lịch sự, lịch thiệp" and they used it to make sentences as in the following examples:

(1) Youngsters should be *polite* to the elderly.
(2) She greeted me with a *polite* smile.
(3) You should wear *polite* clothes to a job interview.

To clarify and explain how to use the word “polite” in these cases, monolingual dictionaries such as the Oxford Learner's Dictionary or the Cambridge Learner's Dictionary are efficient and functional. Examples (1) and (2) are correct because “polite” is used to show good manners or respect for the feelings of others, whilst the learner has wrong word use in example (3) with the collocation "polite clothes". "Formal wear" or "formal clothes" would be appropriate because "formal" can be used in the style of dress, speech, writing, behaviour, etc., and it means “very correct and suitable for official or important occasions". It seems apparent that L1 transfer in the example is negative. According to Feng (2020), transfer is definite and is always negative from L1 to L2, and difficulties in L2 learning can be heralded by L1-L2 discrepancies. In this case, L1 is Vietnamese, and the differences between Vietnamese and English result in a negative L1 transfer. It is noted that most of them have an inability to develop decoding skills. That is the competence to recognize and extract meaningful information about the word from the context, and that results in inaccuracies in word choice and collocation.

Grammar

Hinkel (2004) claims that grammar teaching can be fruitful if it is cumulative. Notably, there are substantial discrepancies between their homework tasks and writing tests in terms of grammatical error-free. In other words, grammatical errors are extremely rare, with the great assistance of grammar check software programs like Grammarly, GrammarCheck, and SpinBot, but they make major grammatical mistakes in their writing tests. These grammar check programs should be viewed as complementary tools in autonomous learning, not an alternative to teaching. These programs are crucial to raise their awareness of grammatical issues and boost their language proficiency with correct sentence input. Nevertheless, when too many grammatical errors occur, they tend to ignore and fail to understand the suggestion of the software. It comes as no surprise to learn that the majority of respondents access free online Google Translate to transfer L1 messages to L2, but the meaning definitely changes from Vietnamese to English, as such technology is able to go beyond a surface level of grammatical accuracy but fails to ensure L2 writing style.

(1) Người ta có xu hướng mua sắm quá mức là do đó là thói quen đến từ việc bị quá nhiều

áp lực và muốn làm một cái gì đó để giải tỏa.

(2) People tend to overshop because it is a habit that comes from being under too much pressure and wanting to do something to relieve it.

(3) Excessive shopping often stems from a habitual response to stress, driven by the desire to alleviate pressure.

(2) translated by Google Translate is grammar-error free, but the language of (2) is not academic enough with numerous redundant words such as "comes from being", and "wanting to do something". That makes (2) lengthy and inappropriate in academic writing. (3) is a suggested correction with better wording. Specifically, student writers tend to compose in L1 and try to translate the text into L2, resulting in grammatical and lexical errors, unidiomatic and strange-sounding prose. (Wang, 2003).

Table 3

The most common grammatical mistakes in their essay writing

Common grammatical mistakes
1. Tenses
2. Prepositions
3. Part of speech
4. V-ing or To-inf
5. Subject and verb agreement
6. Present and past participles
7. Conjunctions
8. Comparisons
9. Relative pronouns
10. Passive voice
11. If conditions

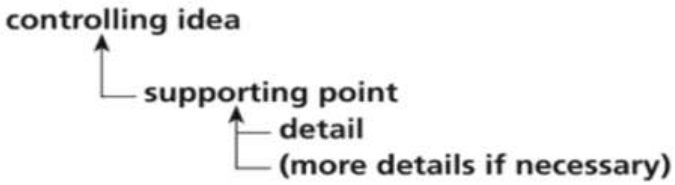
Table 3 outlines the most frequent grammatical mistakes in their essays. By far, the most frequently occurring grammatical mistakes found are tenses (present simple, past simple, and present perfect). The next two common mistakes involve the use of prepositions and parts of speech. Relative pronouns are the least frequent grammatical mistakes produced by the students.

Coherence – Cohesion

Chong and Ye (2020) define coherence as the logical progression of ideas and information, while cohesion refers to the connection of ideas and relationships between sentences" (p.5). In an essay, all ideas should relate to the thesis statement, and the supporting ideas in the main body paragraph should connect with the topic sentence. For coherence in writing, the sentences must bind together; that is, the movements from one sentence to the next counterpart must be logical and smooth. There must be no sudden jump. Some strategies, such as using repetition of key nouns, consistent pronouns, transition signals, and logical order, are introduced and applied to support writer learners.

Figure 1.

A suggested model of a well-written paragraph. (p.11)



The structure of a well-organized paragraph by Chin et al. (2012) is illustrated in Figure 1. Most participants stated that supporting and further explaining controlling was the most demanding and challenging for them because of their poor understanding and limited evidence. Hence, there is an urgent need to help them with this section. The student writers are conventionally introduced to how to write a well-ordered paragraph consisting of a controlling idea in a topic sentence, followed by reasoning and examples or substantial evidence to develop and support the controlling idea.

Research Question 2: What are the experience and real-world knowledge gaps in IELTS writing task 2 for engineering freshmen in DUT-UDN?

Experience and real-world knowledge gaps

In addition to linguistic competence, knowledge of content is also core to the writing teaching process and strategies. Learner writers are expected to express their personal experience and general knowledge on an issue to handle the task properly. The discussion of an issue of general interest and the disclosure of personal thoughts and viewpoints is, of course, a communicative purpose of IELTS writing task 2, and it is widely viewed as an argumentative genre. It should be noted that grammar and lexical correction do not guarantee high-quality academic writing. Writers are expected to reflect their in-depth knowledge and analytical skills on that topic. These expectations can be met by employing conventional supporting techniques such as critical reasoning, typical examples, statistics, and quotations. It would be, therefore, functional and useful for them to work on these supporting techniques. In addition to teaching these techniques, they have more chances to view writing rubrics with the requirements of supporting techniques in their grading.

Unfortunately, all three supporting techniques are absent in the body paragraphs of diagnostic writing tests, but reasoning is primarily exploited to support and develop the controlling ideas. Reasoning is the technique that a writer uses to explain where, how, and why the evidence supports the claim.

Table 4

Average frequency of examples, statistics, and quotations in 202 diagnostic and final term writing tests

Diagnostic writing tests (202)			**Final-term tests (202)**		
Examples	Statistics	Quotations	Examples	Statistics	Quotations
11	0	0	55	15	17

Without the introduction of substantial evidence in their supporting techniques, it can be seen that the frequency of examples, statistics, and quotations in diagnostic writing tests is much less than that of examples, statistics, and quotations in final-term writing tests. Eleven examples are occasionally used in 202 diagnostic writing tests, while statistics and quotations are completely absent. In other words, reasoning is a dominant way to support and elaborate their argument. In contrast, there is a surging use of examples in final-term writing tests, statistics, and quotations, which are nearly equal at 15 and 17, respectively. It can be seen that a product-oriented approach can work well to raise their awareness of knowledge content and narrow the knowledge content gaps.

Typical examples

Personal or impersonal examples

The most common question writer learners ask is whether giving personal or typical examples is appropriate. There is no straightforward answer because they are guided to write about their own experience. Let's consider the following example of their writing.

High-speed internet allows learners to save loading times of websites, videos, and other educational resources, ensuring that students don't waste time waiting for content to load.

Table 5

An example of developing or supporting the main argument

Topic sentence	High-speed internet can provide a significant boost in e-learning.
Reasoning	High-speed internet allows learners to save loading times of websites, videos, and other educational resources, ensuring that students don't waste time waiting for content to load.
Personal example	For example, I often use e-learning platforms to join online courses, download textbooks and materials, and watch recorded videos.
Typical example	Zoom, Ms-team, or Google Meet can serve as typical e-learning platforms, allowing learners to engage globally in virtual classes and interact with their classmates and teachers without technical interruption.

The learner gets their own example to illustrate and develop the topic sentence. A typical

example works better than a personal example in argumentative writing because of its art of persuasion. It is evident that making the writer disappear in giving examples can assure objective and strong argumentation, but avoid a subjective stance. Particularly, the typical example in this case can show their understanding of updated platforms as well as trendy teaching and learning methods.

Increasing cultural awareness and knowledge

According to the British Council, IELTS is widely recognized as an international test in 140 nations and 4,000 locations. Thus, cultural differences should be taken into consideration when giving personal examples in supporting techniques. To specify, when there are cultural differences between IELTS candidates and markers, personal examples are occasionally outside of the IELTS markers' awareness, which can lead to potential sources of misunderstanding. This approach is discussed to enable IELTS candidates to consciously recognize the influence of cultural factors in personal examples in their argumentative essays.

Table 6

An example of personal experience in the supporting idea

Topic sentence	Cultural activities can bring numerous financial gains
Reasoning	Producing a good movie can make a lot of earnings and boost tourism growth.
Example	For instance, the movie “Mai” directed by Tran Thanh can earn over 100 billion VND.

The gross of "Mai", a Vietnamese commercial film released in 2024, is mentioned to illustrate the financial rewards of cultural products. In this case, the concept of this film may be unfamiliar to foreign markets' culture and can lead to misunderstanding due to differences in social knowledge and relevance. Utilizing some internationally recognized highest-grossing movies, such as Titanic (1997) or Avatar (2009), to avoid intercultural communication barriers and guarantee communication across cultures is highly recommended. It is noted that the utilization of common culture or national identity in giving examples is still debated and controversial. The analysis of 202 IELTS writing task 2 samples reveals that Vietnamese past and current socio-economic events, ranging from Vietnamese wars and festivals to contemporary game shows and icons, are available in their examples to support and develop controlling ideas. Therefore, communication across cultures should be considered when giving examples to support and develop controlling ideas.

Sapir argues that "language does not exist apart from culture, that is, from the socially inherited assemblage of practices and beliefs that determines the texture of our lives" (Sapir, 1921, p. 221). Language and culture are closely linked, and cultural practices influence how language is used in different contexts. They should bear in mind that understanding a language's culture is crucial, and language is not merely a system of communication but a medium through which cultural norms, values, and beliefs are taught and reinforced.

Statistics

Unlike examples, statistics are concise, relevant, and authentic with cited sources. Therefore, they are substantial and powerful supporting details, but are occasionally found in their academic essays. Most of the respondents claim that it is too challenging for them because they have to read and remember the right figures and cited sources of statistical data across a wide range of topics. They share a belief that this kind of supporting ideas is highly valuable and convincing because it doesn't work much to persuade readers or markers by using your own ideas. Here's an example of a statistical supporting technique:

(1) According to FlexOS, by 2023, around 28% of employees globally will work part-time or full-time from home. This shows that the way people work is changing for the better.

The above example includes the cited source and time and shows a number, but the percentage "28% of employees globally work from home" cannot speak for itself and needs further explanation and interpretation to obtain the desired outcomes. It is suggested that they exploit such linking language to clarify further and talk about the source as: "the data also demonstrates …"; "this article goes on to say …." and others. Note that all of the 15 statistical data just mention the sources' names and titles of the sources, but the author is responsible for stating the information and data they are citing. For example, instead of mentioning the author of the article, they indirectly quote "According to CNN, …."

The very low frequency of statistical data in their essays (15 cases in 202 essays) somehow reflects the absence of their background statistics and knowledge in this subject. Extensive reading is likely the major source of statistics-focused input to address the lack of statistical evidence. Statistics-focused input relates to the learners' reading and bearing in mind the statistical data, while statistics-focused output involves them using and quoting concise data to make their argument as efficiently as possible.

Quotations

Many quotations are found in articles, reports, advertisements, slogans, and other everyday materials. The question is how and why we quote in academic writing.

From the context-based perspective, a quotation is classified into two categories, namely (i) direct quotations and (ii) indirect quotations**.** Fetzer et al. (2015) view direct quotation as:

> "verbatim speech report or as citation of something which has been said/ written before. Since it represents a discursive excerpt in a verbatim manner, it has generally been considered as non-evaluative use of language. Direct quotation, on the other hand, is defined as a reference to some prior speech report presented and evaluated from the present speaker's perspective, as reflected in deictic shifts". (p.250).

Here are some main rules for direct quotations. Firstly, reporting verbs such as "claim", "argue", "state", and others can appear before, in the middle of, or after borrowed information, and the reporting phrase "according to" can stand before or after but not in the middle. Secondly, including the source of the borrowed information with the reporting expression gives authority to your writing because it allows readers to trust your information. Finally, put quotation marks around information that you copy word for word from a source. Avoid using quotation marks

with paraphrases, summaries, or direct quotations. In contrast, to make indirect quotations grammatically well-formed, a backshift of tenses, time expressions, as well as subject and object changes is required. Here is an example of direct quotations from collected data.

(1) Nelson Mandela said, “Education is the most powerful weapon that you can use to change the world.”

(2) Aristotle's words, "The roots of education are bitter, but the fruit is sweet," reflect the hardships of education.

(3) The former and gifted president Ho Chi Minh once said, "For the benefit of ten years, plant trees. For the benefit of a hundred years, cultivate people."

Direct quotations outnumber indirect quotations in their academic essay, but most are brief and easy to remember due to the time constraint of the IELTS writing test. These three examples include three direct quotes from famous people with three appropriate quotation marks. Nevertheless, most of the quotes are inserted without further explanation to support their controlling ideas. It is, therefore, crucial to remind them to employ quotes to demonstrate and clarify the main ideas, not just to take up space. From a pragmatic perspective, proverbs and quotes of famous or iconic people are ubiquitous and functional in academic citations because they allow writers to persuade and reinforce their argument or statement.

Discussion

The findings of this study reveal that the engineering freshmen at the University of Da Nang use Tier 1 words at much higher frequencies than Tier 2 and 3 words on account of their repetitive use and straightforward and simple meaning. In particular, they have marked preferences for word-by-word translation, which is very old-fashioned and is mostly utilized in their writing. Mother tongue inference is undoubtedly available and nearly unavoidable for limited and modest learners. It is a noticeable phenomenon in the learning of writing by non-native English learners. Dulay et al. (1982:77) view mother tongue inference as "the automatic transfer, due to habit, of the surface structure of the first language into the surface of the target language ". That is to say, the learners recall their first language rules while they express their ideas and thoughts in the target language. There is a considerable similarity between the outcome of this study and that of Nguyen and Nguyen (2022) in terms of vocabulary in writing IELTS task 2. Specifically, 129 out of 205 participants in their study point out that being influenced by Vietnamese is one of the top problems the students have to handle.

In addition to lexical instruction, grammatical features are essential for any L2 teachers and learners in writing because it would be hard to understand their writing if grammatical errors occur with high frequency and impede communication. It is widely agreed that teaching the writing process and discourse and rhetorical features of a specific genre is much time-saving and unchallenging than teaching such language skills as lexis and grammar on account of the unlimited academic words and advanced, complicated, unfamiliar English grammar. To put it plainly, it takes years to master English Tiers 2 and 3 and advanced grammatical structures, and it is demanding to ask them to write academic essays without sufficient linguistic proficiency.

Participants admitted that they had devoted a great deal of time, energy, and effort to learning and mastering English grammar rules, but frequent grammatical mistakes disappoint and demotivate their writing process. Specifically, they explained that they were entirely familiar with some English grammar rules, such as tenses, passive voices, if conditions, and others, but putting these rules in writing is another story. Nghi and Truong (2023) also state that nearly 60 percent of students from grades 6 to 12 at the International Pacific School - Dong Na totally agree that grammar is always a difficult aspect of academic writing. Ferris and Hedgecock (2013) acknowledge and emphasize the significance of error treatment in exposing learners to their inaccuracy in grammar and limited grammatical ranges. Truscott (1999), representing the perspectives of teachers following process-oriented writing theories, argued that correcting errors in a written composition might help learners fix those errors. "Focused" and "Unfocused" correction feedback are discussed and differentiated in error correction. The former selects specific errors, such as tenses, to be corrected and ignores other errors, whilst the latter is involved in correcting all errors in learners' essays. The latter type is also viewed as "extensive" because it works on all existing errors. Extensive correction feedback is highly challenging and time-consuming, but efficient for learners on account of the myriad errors and new approaches to teaching and learning the writing process. That is, there is a shift from putting emphasis on grammar instruction and error correction to focusing on supporting ideas, drafting, revising, and peer reviewing. These changes allow students to get more productive and motivated to pay attention to ideas rather than grammar accuracy. In addition to offering skilled techniques and appropriate strategies for teaching L2 writing, grammar and vocabulary, learners should be guided to build their own thoughtful selection of sets of Tier words 2 and 3 and common grammar rules of formal written English for future use as well as do some pair review to recognize and correct some popular and repetitive lexical and grammatical mistakes in their work to boost substantial improvements for their learning.

Notably, this study presents some experience and real-world knowledge gaps that are not found in any previous studies in the literature review. Numerous advocates, including me, are inclined to use a communication approach to teach writing. Linguistic competence alone cannot guarantee a successful communication attempt. However, cultural knowledge, facts, and experience are paramount. In other words, the writer and reader should share a certain level of general knowledge, practices, beliefs, values, experience, and others to decode and encode the written message. Kramsch (1998) statesLanguage, " is the principle means whereby we conduct our social lives. When it is used in contexts of communication, it is bound up with culture in multiple and complex ways."

In sum, it is apparent that lexical and grammatical knowledge is a primary focus in IELTS writing task 2. Working on that knowledge is beneficial for students' performance in IELTS writing task 2 (Nguyen & Nguyen, 2022; Nghi & Truong, 2023). In addition to lexical and grammatical knowledge, writing improvement in this work concentrates on idea development and evidence. In other words, this study suggests that teaching and learning English composition rely not only on lexical-grammatical range and accuracy but also on content knowledge. It is commonly agreed that good writers are good readers (Fitzgerald &Shanahan, 2000; Langer & Flihan, 2000). To become good writers, students are recommended to read articles, books, and authentic online writing resources to extract the essential information and

evidence for their written composition rather than relying on their own viewpoints and ideas.

Conclusion

The study is hoped to benefit both L2 teachers and learners who aspire to make desirable progress in academic writing. In conjunction with teaching the writing process, discourse, and rhetorical features of text, it is essential to teach language skills that L2 writers need to succeed in mainstream university classes. The study with 202 collected short essays attempts to redress this perceived shortcoming by identifying and describing the formal syntactic and lexical characteristics of academic text and advocating the explicit teaching of these key structures and lexical chunks as well as put an emphasis on presenting evidence to support and develop the controlling ideas by teaching skills to argue as well as reason critically, give solid examples, cite statistical evidence. The awareness of communication across cultures in learning and teaching academic writing should be raised. It is evident that bridging the language, experience, and real-world knowledge gaps can greatly impact the test-takers' performance in IELTS writing task 2. Due to time and space constraints, pair reviews and feedback are not discussed, while the author tried to present a holistic approach to teaching and learning academic writing.

Acknowledgments

I would like to thank all the participants and reviewers for their contributions, comments, and suggestions on the drafts of this article.

References

Badger, R. (2024). *Teaching and Learning the English Language: A Problem-Solving Approach.* London: Bloomsbury Publishing Plc.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing Words to Life: Robust Vocabulary Instruction.* New York: Guilford Press Book/Childcraft International.

British Council. (n.d). *IELTS Task 2 Writing Band Descriptors (public version)*. Retrieved from https://takeIELTS.britishcouncil.org/sites/default/files/IELTS_writing_band_descriptors.pdf

Chin, P., Koizumi, Y., Reid, S., Wray, S., Yamazaki, Y. (2012). *Academic Writing Skill 1 Student's Book*. Singapore: Cambridge University Press.

Chong, S. W., & Ye, X. (2020). *Developing writing skills for IELTS: A research-based approach*. Oxon: Routledge.

Davies A. (2008). Assessing academic English. Testing English proficiency 1950—1989: The IELTS solution (*Studies in Language Testing, 23*). Cambridge University Press.

Dulay, H. C., Burt, M. K., Krashen, S. (1982). *Language two*. New York: Oxford University Press.

Evans, M. (2019). A Study of Academic Genre: Exploring Writing in English for University Purposes with Vietnamese Students. *VNU Journal of Foreign Studies, 35*(1), 63-68.

Fan, M.Y. (2000). The Dictionary Look-Up Behavior of Hong Kong Students: A Large Scale Survey. *Education Journal, 28(1),* 123-138.

Feng, H. (2020). *Form, Meaning and Function in Collocation: A Corpus Study on Commercial Chinese-to-English Translation*. New York, NY: Routledge.

Ferris, D. (2011). *Treatment of Error in Second Language Student Writing*, Second Edition. Michigan: The University of Michigan Press.

Ferris, D. & Hedgecock, J. (2013). *Teaching L2 Composition: Purpose, Process, and Practice*. New York, NY: Routledge.

Fetzer, A, Berlin, L., Weizman, E. and Berlin, L. (2015.) *The dynamics of political discourse: Forms and functions of follow-ups.* Amsterdam: John Benjamins

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35(1)*, 39–50. https://doi.org/10.1207/S15326985EP3501_5

Hamp-Lyons. (1991). *Assessing Second Language Writing in Academic Contexts*/ edited by Liz Hamp-Lyons. N.J.: Ablex, 1991

Hinkel, E. (2004). *Teaching Academic Esl Writing: Practical Techniques in Vocabulary and Grammar.* New York. Routledge

Kramsch, C. (1998). *Language and culture*. Oxford: Oxford University Press.

Langer, J. A., & Flihan, S. (2000). Writing and reading relationships: Constructive tasks. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing: Research, theory, and practice* (pp. 112–139). International Reading Association. https://doi.org/10.1598/0872072681.5

Milton, J., Wade, J., & Hopkins, N. (2010). Aural Word Recognition and Oral Competence in English as a Foreign Language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. M. Torreblanca-López (Eds.), *Insights into Non-Native Vocabulary Teaching and Learning* (pp.83-98). Multilingual Matters.

Nghi, N. P. A., & Truong, T. N. (2023). Difficulties of Vietnamese Students in Learning Academic Writing. *International Journal of Asian Education, 4(2),* 106–114. https://doi.org/10.46966/ijae.v4i2.340

Nguyen, H.N. & Nguyen, D.K. (2022). Vietnamese Learner's Performance in The IELTS Writing Task 2: Problems, Causes and Suggestions. *International Journal of TESOL & Education, 2(1)*, 170-189. DOI: https://doi.org/10.54855/ijte.222111.

Read, J.(2015). *Assessing English proficiency for university study. Basingstoke*, UK: Palgrave Macmillan.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World Inc

Schnoor, M. (2003). *Different Types of Dictionaries.* Germany: Grin Verlag Publisher.

Taylor L., Falvey P. (Eds.). (2007). IELTS collected papers: Research in speaking and writing assessment (*Studies in Language Testing, 19*). Cambrige: Cambridge University Press.

Truscott, J. (1999). The Case for "the Case for Grammar Correction in L2 Writing Classes": A Response to Ferris. *Journal of Second Language Writing, 8(2)*, 111-122. https://doi.org/10.1016/S1060-3743(99)80124-6

Tran, T. T., & Truong, D. M. D. (2021). The Influence of Single-Clause Sentences on IELTS Writing Task 2 Band Score. *International Journal of TESOL & Education, 2(1)*, 23–39. https://doi.org/10.54855/ijte2202012

Wang, C. M. (2003). The compensation hypothesis in L2 learning. *Foreign Language Research, 112(1)*, 1-5.

Weir C. J., O'Sullivan B. (2017). *Assessing English on the global stage: The British Council and English language testing*, 1941—2016. Equinox.

Biodata

Dr. Nguyen Thi Tu Trinh is an English lecturer of University of Science and Technology – The University of Da Nang, Vietnam. She won an Erasmus+ Mundus Fellowship in Université Côte D'azur in Nice, France and joined a visiting scholar program in Southern Connecticut State University in 2019. Her teaching focuses on areas of English teaching and communication skills. Her research interests have involved linguistic theory; its application to classroom teaching; Functional Grammar and Technology-assisted teaching (Web-based learning) to boost English learning.

The Benefits and Challenges of AI Translation Tools in Translation Education at the Tertiary Level: A Systematic Review

Nguyen Thi Nhu Ngoc[1*], Tran Thanh Truc[1], Nguyen Ngoc Hoang An[1], Lam Hoang Phat[1], Nguyen Hua My San[1], Tran Nguyen Anh Thu[1]

[1]University of Social Sciences & Humanities, Vietnam National University Ho Chi Minh City, Vietnam
[*]Corresponding author's email: nhungoc@hcmussh.edu.vn
[*]https://orcid.org/0000-0002-5015-2841
https://doi.org/10.54855/ijte.25527

Received: 25/01/2025 Revision: 14/05/2025 Accepted: 20/05/2025 Online: 20/05/2025

ABSTRACT

The growth of AI-powered technologies has highlighted transformative impacts. This study presents a systematic review of the benefits and challenges of AI translation tools in tertiary-level translation education. It aims to identify how they support and hinder students and teachers. The data were extracted from 20 peer-reviewed articles (2014-2024) on six academic databases, using standardization and thematic synthesis based on the PRISMA guidelines. The qualitative findings revealed eight core benefits: enhanced translation efficiency, improved vocabulary and grammar, post-editing support, increased learner motivation, professional and technical preparedness, accessibility and inclusion, reflective learning and personalization, and teacher support and pedagogical innovation; and seven key challenges: overreliance on AI, contextual inaccuracies, digital inequity, insufficient training and pedagogical gaps, ethical and privacy concerns, usability issues for senior lecturers, and lack of institutional support and curriculum integration. Valuable insights and recommendations were then offered to refine translation pedagogy with effective AI tool integration.

Keywords: AI translation tools, benefits, challenges, translation education, tertiary level

Introduction

The 21st century has seen remarkable technological advancements. Thus, manual translation alone is insufficient for the demands of the field, necessitating the integration of artificial intelligence (AI) translation tools to streamline processes and enhance translation quality. AI-powered translation technologies have had a significant impact on the industry, extending their influence beyond professional practice to education (Koka, 2024). Digital technologies have

CITATION | Nguyen, T. N. N., Tran, T. T., Nguyen, N. H. A., Lam, H. P., Nguyen, H. M. S., & Tran, N. A. T. (2025). The Benefits and Challenges of AI Translation Tools in Translation Education at the Tertiary Level: A Systematic Review. *International Journal of TESOL & Education*, *5*(2), 132-148. DOI: https://doi.org/10.54855/ijte.25527

transformed daily life, affecting how individuals seek information, communicate, and learn in education (Chassignol et al., 2018). Thus, translation education is no exception.

Computer-assisted translation (CAT) tools are important in improving the translation process and quality, which is helpful in training competent translators (Omar et al., 2020). Currently, in this AI-driven time, teachers should be reliable guides for their students in addition to their traditional role of language instructors (Phuong, 2024). Essentially, the role of AI translation tools in designing curricula needs to be highlighted in the combination of human translation skills with AI-assisted tasks. Thus, students are required to enhance both linguistic competence and technological competence (Tavares et al., 2023). Once the translation industry incorporates AI and CAT tools as a must, university programs should provide their students with an effective use of AI in their translation practice and promote their critical thinking and ethical standards (Hazaea & Qassem, 2024). The gap between the academic translation environment and the professional one will be narrowed with an understanding of the AI role (Tian, 2024).

As a result, a critical comparison and analysis of the relevant existing studies will provide useful and evidence-based insights. Thus, this study is a systematic review of the benefits and challenges of the use of AI translation tools in tertiary-level translation education. Its aim is to examine their support for translation training and identify difficulties faced by both students and teachers. The results provide students, teachers, and translation program developers with insights to refine translation curricula and technology integration strategies, offering implications for translation pedagogy and student learning outcomes.

Literature Review

Current Trends of Using AI Translation Tools in Translation Education

Alharbi (2023) demonstrates that new AI technology has helped overcome language and cultural barriers through rapid translations. Therefore, many high-technology inventions and AI advancements have become integral to our lives, assisting us in numerous ways, especially in the translation industry. AI tools, such as chatbots and automated assistants, support the translation process by facilitating language processes and promoting diverse cultural communication. Recognizing that AI translation tools have indirectly aided the translation industry, Bates et al. (2023) highlight the vital role of AI, specifically in bridging training gaps in academic settings.

AI technology has transformed translation education by streamlining the learning process through the integration of AI translation tools, providing real-time assistance for both teachers and students. These tools not only promote language acquisition but also provide professional training (Alharbi, 2023). The increasing demands of technology for translation practice mean that learners' familiarity with these tools helps them be better prepared for the job market. Besides providing technical support accompanied by real-time feedback, detecting errors, and suggesting phrases, the tools help them understand language in numerous settings, thereby enhancing their translation quality (Al-Rumaih, 2021; Bakhov et al., 2024; Deng & Yu, 2022; Han, 2020; Liu & Afzaal, 2021; Koka, 2024; Odacıoğlu & Kokturk, 2015; Omar & Salih, 2024; Tavares et al., 2023; Zhang, 2023).

Therefore, AI technology has become part of translation education. It helps break language and cultural barriers with time-saving and productive tools. As research shows, AI translation tools are useful aids to speed up translation and good preparation for students to meet the industry's requirements, though there exist some hindrances. The exploration and implementation of AI technology in translation education at the tertiary level are encouraged.

Research Questions

To achieve the research's aim and objectives, this systematic review is conducted to address the two research questions as follows:

1. What benefits do students and teachers get from using AI translation tools?
2. What challenges do students and teachers face in using AI translation tools?

Methods

Design of the Study

The study employs a systematic review methodology. It involves a structured process of identifying, selecting, and critically assessing relevant studies, followed by a thematic synthesis of the findings (Petticrew & Roberts, 2006). In this review, specific inclusion and exclusion criteria were applied to screen 20 empirical studies on the use of AI in translation education. This research design is useful for synthesizing various findings from such interdisciplinary fields as education, linguistics, and technology (Petticrew & Roberts, 2006). It is also an appropriate method to ensure a structured, transparent, and replicable process for identifying, evaluating, and summarizing relevant studies (Tavares et al., 2023); and to adapt to the varied reception in classroom environments (Booth et al., 2016; Gough et al., 2017). Thus, in the case of using AI translation tools in various university contexts, this design is applicable. This review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al., 2021) to set up well-defined inclusion and exclusion criteria and systematic database search strategies.

Pedagogical Setting & Sampling

We selected twenty studies from peer-reviewed journals published in the 2014-2024 period. Data extraction was performed using a standardized form and analyzed through thematic synthesis. Following PRISMA guidelines enhances the reliability, reproducibility, and academic integrity of the review (Moher et al., 2009). In detail, these studies were from the six academic databases, namely Google Scholar, Scopus, Web of Science, ResearchGate, LinguistList, and ERIC, which cover interdisciplinary research, particularly in artificial intelligence, education technology, and translation studies. We combined keywords and Boolean operators for data search. That means we used simple words (AND and OR) to refine searches for more accurate and efficient information. The search string below was used: *(AI-driven translation OR machine translation OR AI-assisted translation) AND (tertiary education OR university education OR higher education) AND (benefits OR challenges OR limitations) AND (translation practice OR translation learning OR translation education).*

To maintain PRISMA guidelines and relevance, specific inclusion and exclusion criteria were established to ensure that updated and contextually relevant studies were included in this systematic review.

Table 1

Criteria for sample selection

Criteria	Inclusion	Exclusion
Study type	Peer-reviewed journal articles, systematic reviews, and empirical studies	Editorials, blogs, opinions
Publication date	Last ten years (2014 - 2024)	Older studies, unless highly relevant
Language	English (or relevant translations)	Non-English without available translation
Participants	translation students and teachers	Professional translators, high school students

This output provided a total of 95 records. After removing 30 duplicates, 65 records remained for title and abstract screening. Six research members were involved in this data search procedure. Two independent reviewer groups (i.e., each has two members) screened all records based on predefined inclusion and exclusion criteria, which considered the study's relevance to AI-driven translation tools in tertiary settings, empirical methodology, and language of publication (English only). After screening and full-text analysis, 20 studies met all inclusion criteria and were included in the final synthesis. Disagreements between the reviewers during the screening or inclusion stages were resolved through discussion. When necessary, a third reviewer group (i.e., the final two members) mediated to reach consensus.

These 20 studies were selected based on their relevance to AI-assisted translation within tertiary education. Papers that focused on the use of AI translation tools in tertiary-level translation training were prioritized. Those in professional or industry contexts were excluded unless they offered pedagogical insights pertinent to academic learning environments. A summary of the studies for the systematic review is presented in Table 2.

Table 2

List of studies for the systematic review

1. Al-Rumaih (2021)	8. Koka (2024)	15. Omar et al. (2020)
2. Bakhov et al. (2024)	9. Laksana and Komara (2024)	16. Roza and Zulhirawati (2023)
3. Bao and Thi (2021)	10. Le and Dao (2019)	
4. Bouguesmia (2020)	11. Li et al. (2024)	17. Soysal (2023)
5. Deng and Yu (2022)	12. Liu and Afzaal (2021)	18. Tavares et al. (2023)
6. Han (2020)	13. Odacıoğlu and Kokturk (2015)	19. Yanti and Meka (2019)
7. Kenny and Doherty (2014)	14. Omar and Salih (2024)	20. Zhang (2023)

Data Collection and Analysis

To address the two research questions, we developed a data extraction framework for consistency and depth in analyzing the selected studies. A list of comprehensive data extraction was utilized to capture essential study details, including the title, authors, year of publication,

research type (qualitative, quantitative, mixed-methods, or review), AI translation tools examined (e.g., Google Translate, DeepL, ChatGPT, SDL Trados, etc.), key benefits and challenges identified, and conclusions relevant to tertiary-level education. This framework enabled us to systematically organize and compare findings across diverse studies, ensuring that both contextual and methodological variables were considered during the analysis.

For data interpretation, we applied a thematic synthesis approach. This qualitative method enabled the identification and comparison of recurring themes related to the two core research questions: the benefits of AI translation tools and the challenges faced by students and/or teachers. In detail, themes stemmed inductively. Data from the selected studies were coded based on their functional role in translation education. Themes were categorized into two main clusters: (1) *benefits* (e.g., efficiency and speed increase, vocabulary and grammar enhancement), (2) *challenges* (e.g., overreliance, accuracy limitations). Due to methodological heterogeneity across studies, a full meta-analysis was not feasible. As a result, narrative synthesis was prioritized to provide a comprehensive, context-rich understanding of the current academic landscape. The summary of the findings can be found in this link: https://drive.google.com/file/d/14rCIZj5dBkavtTi_l4EpkiIgMiMkiIkD/view?usp=drive_link.

The extracted data, including study aims, participants, and reported benefits and challenges of AI translation tools, were organized and managed using Microsoft Excel. Two research groups (each has two members) independently conducted an inductive thematic analysis following Braun and Clarke's (2006) framework. Initial coding was performed separately, and the resulting codes were compared and discussed to develop a shared codebook. Disagreements in coding were resolved through discussion, with a third research group (also consisting of two members) mediating unresolved cases to ensure inter-coder reliability. Final themes were refined through iterative review and validation to ensure consistency and accuracy across the dataset. This collaborative and transparent process helped ensure the reliability and validity of the identified themes.

Results

The Benefits of AI Translation Tools in Tertiary-Level Translation Education

Eight major benefits are identified from the use of AI tools in tertiary-level translation education, as presented in Table 3, including: student efficiency and speed increase (found in 9/20 studies), vocabulary and grammar enhancement (7/20), drafting and post-editing support (5/20), motivation and confidence boost (5/20), professional and technical preparedness (5/20), accessibility and inclusion (6/20), reflective learning and personalization (4/20), and teacher support and pedagogical innovation (3/20). Some common AI tools, such as Google Translate, DeepL, and SDL Trados, have been shown to improve students' translation quality in terms of lexical, syntactic, and semantic aspects. They also enhance students' reflective translation practices through iterative editing, prepare students' professional and technical competencies, and encourage teachers' pedagogical innovations.

Table 3

The benefits of AI translation tools in translation education

Ord.	Benefit	Description	Supporting studies
1	Efficiency and speed increase	AI tools accelerate translation tasks, especially for technical or repetitive content.	Odacıoğlu and Kokturk (2015), Han (2020), Bouguesmia (2020), Liu and Afzaal (2021), Al-Rumaih (2021), Deng and Yu (2022), Tavares et al. (2023), Zhang (2023), Koka (2024), Bakhov et al. (2024) ***(9 studies)***
2	Vocabulary and grammar enhancement	Tools like DeepL, Google Translate, and ChatGPT expose students to new vocabulary, synonyms, and grammar structures.	Yanti and Meka (2019), Le and Dao (2019), Bao and Thi (2021), Deng and Yu (2022), Zhang (2023), Roza and Zulhirawati (2023), Laksana and Komara (2024) ***(7 studies)***
3	Drafting and post-editing support	AI-generated drafts allow learners to revise and reflect, improving translation competence.	Bao and Thi (2021), Tavares et al. (2023), Zhang (2023), Omar and Salih (2024), Bakhov et al. (2024) ***(5 studies)***
4	Motivation and confidence boost	AI tools promote student engagement and increase confidence in translation tasks.	Omar et al. (2020), Bao and Thi (2021), Roza and Zulhirawati (2023), Bakhov et al. (2024), Li et al. (2024) ***(5 studies)***
5	Professional and technical preparedness	Exposure to CAT tools aligns students with real-world market expectations.	Odacıoğlu and Kokturk (2015), Bouguesmia (2020), Han (2020), Omar et al. (2020), Al-Rumaih (2021) ***(5 studies)***
6	Accessibility and inclusion	AI tools are often free, mobile-friendly, and usable offline, aiding students in diverse contexts.	Le and Dao (2019), Yanti and Meka (2019), Bouguesmia (2020), Deng and Yu (2022), Laksana and Komara (2024), Li et al. (2024) ***(6 studies)***
7	Reflective learning and personalization	Students receive tailored feedback, fostering critical thinking and reflective practice.	Soysal (2023), Bakhov et al. (2024), Li et al. (2024), Omar and Salih (2024) ***(4 studies)***
8	Teacher support and pedagogical innovation	Older educators found AI helpful for modernizing pedagogy and offering real-time feedback.	Liu and Afzaal (2021), Soysal (2023), Koka (2024) ***(3 studies)***

These benefits are evidence for enhancing both the teaching and learning experiences. These technologies improve efficiency and speed, particularly in repetitive or technical tasks, while also enriching students' vocabulary and grammar through exposure to diverse language structures. Learners benefit from practical support in drafting and post-editing, which fosters deeper engagement and reflective learning. Additionally, AI enhances motivation, confidence, and professional readiness, ensuring alignment between training in universities' academic environments and real-world expectations in the translation industry. Importantly, these tools promote accessibility and inclusion, providing personalized feedback that enhances critical thinking. Furthermore, teachers find AI invaluable for innovating pedagogy and delivering real-time support.

The Challenges of AI Translation Tools in Tertiary-Level Translation Education

Table 4

The challenges of AI translation tools in translation education

Ord.	Theme	Key Description	Supporting Studies
1	Overreliance and critical thinking decline	Students may become too dependent on AI, affecting creativity, analytical skills, and autonomy.	Yanti and Meka (2019), Bouguesmia (2020), Han (2020), Deng and Yu (2022), Bao and Thi (2021), Tavares et al. (2023), Zhang (2023), Roza and Zulhirawati (2023), Soysal (2023) ***(9 studies)***
2	Accuracy and contextual limitations	AI tools often misinterpret idioms, cultural nuances, and domain-specific content.	Bouguesmia (2020), Liu and Afzaal (2021), Tavares et al. (2023), Zhang (2023), Soysal (2023), Laksana and Komara (2024), Omar and Salih (2024) ***(7 studies)***
3	Digital divide and tool accessibility	Inequality in access to devices, reliable internet, and paid platforms hinders full integration.	Odacıoğlu and Kokturk (2015), Al-Rumaih (2021), Koka (2024), Li et al. (2024) ***(4 studies)***
4	Insufficient training and pedagogical gaps	Both educators and students lack structured training in AI and post-editing, causing skill gaps.	Kenny and Doherty (2014). Odacıoğlu and Kokturk (2015), Omar et al. (2020), Al-Rumaih (2021), Omar and Salih (2024) ***(5 studies)***
5	Ethical and privacy concerns	Using AI tools may expose sensitive data and raise concerns over data ownership and job loss.	Kenny and Doherty (2014), Bouguesmia (2020), Liu and Afzaal (2021), Soysal (2023), Koka (2024), Omar and Salih (2024) ***(6 studies)***
6	Usability issues for senior educators	Older lecturers struggle with non-intuitive interfaces, affecting adoption.	Bouguesmia (2020), Soysal (2023), Koka (2024), Li et al. (2024) ***(4 studies)***
7	Lack of institutional support and curriculum integration	Few programs offer comprehensive AI or CAT tool training; many syllabi are outdated.	Kenny and Doherty (2014), Odacıoğlu and Kokturk (2015), Han (2020), Omar et al. (2020), Al-Rumaih (2021) ***(5 studies)***

As highlighted in Table 4, the increasing incorporation of AI in translation education brings with it seven significant challenges at the university level, including: overreliance and critical thinking decline (found in 9/20 studies), accuracy and contextual limitations (7/20), digital divide and tool accessibility (4/20), insufficient training and pedagogical gaps (6/20), ethical and privacy concerns (5/20), usability issues for senior educators (4/20), lack of institutional support and curriculum integration (5/20). The studies show that many students face overreliance on AI outputs, limited contextual accuracy, and digital inequity. Both teachers and students lack structured training. Moreover, ethical issues such as data privacy, authorship, and the professional displacement of human translators require urgent attention.

As seen in the findings, the foremost concern is the potential for students to become overly reliant on AI, which can affect the development of independent critical thinking and creativity.

Furthermore, AI's limitations in interpreting nuanced language, idioms, and cultural contexts can compromise accuracy. Structural issues, such as unequal access to devices and internet connectivity, contribute to digital divides. Additionally, both teachers and students frequently lack formal training in AI and post-editing practices. Ethical and privacy issues also lead to worries about owning data and losing jobs. Senior lecturers find it challenging to use AI tools and get inadequate institutional support, leading to outdated curricula without sufficient integration of AI tools.

Discussion

Based on the thematic analysis of the 20 studies, we could easily identify that a key benefit of AI translation tools (especially MT and CAT) is the acceleration of the translation process. Both students and teachers agreed that AI aids them in completing tasks faster, freeing up time for post-editing, critical reflection, or supplementary tasks. Bouguesmia (2020) and Tavares et al. (2023) revealed that students appreciated NMT tools for speeding up draft creation, which allowed more time for revision. Han (2020) emphasized that CAT tools like TM significantly reduce redundancy, making translation faster. Also, Zhang (2023) and Deng & Yu (2022) affirmed that MT tools help students manage complex texts, reducing cognitive load. Omar et al. (2020) found that students' translation production was faster when using SDL Trados and other CAT tools. Therefore, efficiency is a primary benefit of AI translation tools, aligning with classroom goals and industry standards and reflecting the views of students and teachers in various educational contexts.

As for vocabulary and grammar enhancement, AI translation tools often suggest synonyms, grammatical structures, and terminology, helping students to expand their linguistic awareness, especially when comparing AI output with manual solutions. Yanti and Meka (2019) found that students learned new vocabulary incidentally through app usage. Le and Dao (2019) observed that students exploited online dictionaries to gain a nuanced understanding of word usage and collocations. Bao and Thi (2021) revealed that online dictionaries and grammar checkers (e.g., Grammarly, Longman Dictionary) helped students revise their language more effectively. To be more specific, Laksana and Komara (2024) showed that DeepL improved students' grammar awareness and vocabulary learning. Roza and Zulhirawati (2023) affirmed that ChatGPT enhanced vocabulary acquisition, with students noting clear improvement in word choice. It can be seen that AI tools function as both translation engines and incidental learning environments, offering instant vocabulary and syntax feedback that aids language acquisition.

AI tools show their great support for translation drafting and post-editing. Many of the 20 studies found that students often use AI-generated translations as first drafts, which they then post-edit. This process improves linguistic accuracy and helps them practice industry-relevant skills like MTPE (Machine Translation Post-Editing). In Tavares et al. (2023), students could

use MT tools to generate and refine their translation drafts effectively. Omar and Salih (2024) stated that translation quality was enhanced in the post-editing thanks to AI integrating into classroom practice. Zhang (2023) and Bakhov et al. (2024) found that AI-assisted post-editing encouraged critical reflection and iterative improvement. In brief, the use of AI as a drafting tool supports both learning and skill-building. When combined with post-editing, it fosters reflective and quality-focused translation pedagogy.

Motivation and confidence boost are clearly evidenced from these studies. AI tools reduce intimidation for would-be translators at universities, giving them a tangible starting point. This increases motivation and reduces the fear of failure, especially with difficult or technical content. Bakhov et al. (2024) demonstrated improved student motivation and participation after integrating AI-assisted tools. Omar et al. (2020) reported a notable rise in student confidence when working with CAT tools. Li et al. (2024) found that perceived enjoyment and usability directly impacted students' intention to adopt AI tools in future learning. Obviously, motivational benefits are a critical pedagogical strength of AI tools, especially for underconfident or early-stage translators.

AI tools are a good means for students' professional and technical preparedness. In detail, familiarity with CAT tools and AI-based systems prepares students for real-world professional translation environments, especially in localization, legal, and technical fields. Han (2020) and Odacıoğlu and Kokturk (2015) argue that industry-standard tools (e.g., SDL Trados) must be taught in academia to meet market demands. Omar et al. (2020) found that exposure to software improved student readiness for specialized domains. Al-Rumaih (2021) linked CAT tool proficiency to improved job prospects. Thus, AI tool training is not only a pedagogical strategy but also a career-readiness imperative.

AI-driven tools provide accessibility and inclusion for translation practice. Their features, such as being mobile and free, offer equalizing potential for students with limited resources or access to traditional classroom infrastructure. Le & Dao (2019) and Yanti and Meka (2019) emphasized how mobile dictionary apps enable ubiquitous learning. Deng and Yu (2022) found that multimodal AI tools (voice, OCR, etc.) enhanced access across diverse linguistic backgrounds. Laksana and Komara (2024) noted that DeepL offered a user-friendly and inclusive interface for multilingual translation needs. As a result, it is recognizable that translation tools provide low-barrier, high-impact access to translation practice, supporting inclusion across geographies and educational systems.

AI tools serve as valuable resources for reflective learning and personalization by providing instant feedback, which enables students to critically assess and iteratively refine their translations. This approach is consistent with learner-centered, reflective pedagogies. Research by Bakhov et al. (2024) indicated that students became more reflective through cycles of AI-based feedback. Additionally, Omar and Salih (2024) highlighted the significance of the MTPE in fostering higher-order translation reasoning, while Soysal (2023) demonstrated that AI facilitates engagement with complex tasks such as corpus curation and contextual adaptation. Therefore, using AI tools with thoughtful post-editing strategies significantly enhances reflective and personalized learning experiences.

Lastly, AI translation tools are utilized to support teachers and foster pedagogical innovation. They assist teachers in modernizing curricula by providing simulation-based tasks and real-time assessments and by aligning with industry practices. Koka (2024) observed that while some older lecturers faced challenges with technology, many found AI to be advantageous in their teaching once they received proper training. Liu and Afzaal (2021) highlighted how AI transforms traditional lecture models into interactive learning systems. Soysal (2023) advocated for collaborative pedagogies that merge AI with human reflection. Therefore, AI plays a crucial role in pedagogical renewal, particularly when paired with professional development and curricular restructuring.

As synthesized from 20 studies, a consistent challenge identified is students' overreliance on MT outputs, which leads to a decline in critical thinking skills. Research indicates that many students become excessively dependent on these technologies, thereby weakening their independent decision-making, creativity, and analytical reasoning, which are essential skills for professional translation. In Yanti and Meka (2019), many students felt that they became "lazy" due to using Google Translate too much. Similarly, in Bao and Thi (2021), students' heavy dependence on Google Translate without proofreading led to literal or awkward translations. The findings from Tavares et al. (2023) showed that the dependence on MT limits students' skill development. With the same view, Zhang (2023) found that despite the convenience of MT, many students feared that it lessened their creativity and critical thinking. Han (2020) cautioned that repeated reliance on CAT tools could turn translation students into "editors of machines" rather than autonomous thinkers. Therefore, AI overreliance may leave long-term risks because students may move from an active learning state to a passive one.

Notably, accuracy and contextual limitations are another concern. AI tools still do not provide excellent translation equivalents in many cases of cultural awareness, idiomatic nuance, or subject-matter specificity. This is considered a major barrier when students tend to believe in the quality of AI translation suggestions. For example, in Tavares et al. (2023), consistent errors were found in technical domains when students used MT. Meanwhile, Bouguesmia (2020), Laksana and Komara (2024), and Roza and Zulhirawati (2023) reported concerns about their students' inability to handle abstract or culturally embedded phrases using DeepL and ChatGPT. Also, Omar and Salih (2024) showed that many students could not provide accurate equivalents for legal and dialectal nuances in their MT outputs. Soysal (2023) pointed out the limitations of AI's emotional intelligence, particularly in sensitive texts in their students' translation practice. All of these arguments show that AI translation tools lack a "human filter", a component that is essential for faithful, idiomatic, and context-aware translations. This is extremely important for student translators.

There exists some degree of digital divide in tool accessibility. In fact, equitable access to AI tools is still limited by geography, internet infrastructure, and technical fluency, especially among older educators and students in rural or low-income regions. Koka (2024) found that many older lecturers were in a struggle with AI tools due to their low tech literacy and inadequate training. Also, Li et al. (2024) discovered a rural student demographic with limited

exposure to translation technologies, despite general computer literacy. In former studies like Odacıoğlu and Kokturk (2015) and Al-Rumaih (2021), institutional inequities in tool licensing and lab access were revealed. It can be seen that AI transformation at various universities has often been unevenly conducted, which requires a long run to deal with.

Moreover, there are insufficient pedagogical gaps among universities. A major systemic issue is that AI and CAT tools are under-integrated into curricula, and both students and teachers are not always provided with formal training. This leads to misuse, underutilization, and anxiety toward AI adoption. For example, Omar et al. (2020) found that most teachers in their research sample had no formal training in CAT tools. Kenny & Doherty (2014) pointed out that many students and teachers found AI translation tools complex and hard to learn without a deep technical understanding. Al-Rumaih (2021) provided another reason that CAT tools were usually taught in some class meetings of a single course in most universities, leaving students ill-prepared for workplace use. Obviously, with inadequate pedagogical support for AI tool literacy, even the best technologies remain inactive in the classroom, which is challenging in the revision of tertiary-level translation programs.

Ethical and privacy concerns are increasing with the use of AI translation tools. Both students and teachers have now been warned of the risks associated with uploading data to AI platforms, especially in domains like legal or medical translation. Moreover, AI threatens to disrupt translation labor markets, sparking concerns about job displacement. In Koka (2024), ethical issues were raised regarding sensitive data shared on third-party platforms. Soysal (2023) pointed out that there were no privacy protocols and ways to prevent potential biases in AI systems. Omar and Salih (2024) found that MTPE literature has still ignored social and ethical implications, focusing too much on technical efficiency. Supportingly, Kenny and Doherty (2014) raised concerns about ownership and copyright issues in using cloud-based MT systems. Obviously, the AI integration is a certain trend in the world, but ethical fluency has not been considered part of translation education.

Senior lecturers' hesitation to use AI translation tools is of great concern. While their students may adapt quickly to AI technology, many older lecturers haven't refused to have some experience. This may result in their resistance, avoidance, and unequal adoption. Koka (2024) noted that many older lecturers avoid them due to unfamiliarity despite seeing the value of AI tools. Soysal (2023) affirmed that without adequate interface design, adoption lags among less tech-savvy teachers and students. Thus, for experienced teachers in translation programs, it is quite critical.

Lack of institutional support and curriculum integration is an interrelated issue with the challenge above. In translation programs of many universities, CAT or MT are considered elective or isolated components, rather than embedded in their curricula. This weakens their relevance and students' ability to explore AI tool literacy. Han (2020) and Odacıoğlu & Kokturk (2015) urged universities to align curricula with industry standards, including CAT tools and localization platforms. Omar et al. (2020) showed that most universities in their country didn't offer consistent tool exposure. Al-Rumaih (2021) emphasized copyright issues and faculty reluctance due to their universities' lack of investment. Thus, it still takes a long time to reform the use of AI tools as part of core pedagogical methods in translation programs.

The thematic analysis and discussion reveal multiple converging and diverging themes under benefits and challenges. The use of AI translation tools in tertiary-level translation training reveals a complex interplay between technological affordances and pedagogical limitations for both students and teachers. These detailed analyses make it clear that AI tools are both enabling and challenging the evolution of translation education. The focal point of this systematic review is not whether to use AI translation tools due to their benefits and/or challenges, but how to be fully aware of their effects and use them critically, ethically, and inclusively.

Conclusion

The integration of AI translation tools into translation education signifies a notable development in both pedagogy and technology. This systematic review synthesizes findings from 20 studies, emphasizing eight major benefits of AI translation tools in transformative potentials and seven challenges associated with the use of AI-assisted translation in tertiary-level training environments.

Based on the thematic findings of this systematic review, the following recommendations are proposed:

(1) *Curriculum redesign for the use of AI translation tools*: Translation programs should move beyond elective or standalone courses and embed AI-driven tool training across the curriculum. This includes the teaching of MTPE, terminology management, and ethical translation practices. It is essential to reflect the digital demands of the Industry 4.0 translation market in translation training (Han, 2020).

(2) *Critical and reflective use of AI translation tools*: Students should be taught to use AI tools as assistants, not authorities. Teachers should prioritize critical evaluation, post-editing, and comparison tasks that strengthen human judgment and discourage overreliance. Students benefit from AI-generated drafts but need training to refine and critique them effectively (Tavares et al., 2023).

(3) *Training and support for teachers*: Professional development programs must be offered to equip teachers, especially senior ones, with the digital fluency needed to adopt AI tools confidently. The lack of technical confidence among older instructors is a barrier to AI adoption in classrooms (Koka, 2024).

(4) *Ethical guidelines and AI literacy*: Universities should create clear ethical policies surrounding the use of AI in translation education as part of the AI-use policies for their whole systems. Students need early exposure to issues like data privacy, authorship, and AI bias to become responsible digital translators. Integrating ethical awareness into AI-focused translation education helps address the risks of bias and misuse (Soysal, 2023).

(5) *Invest in infrastructure and access equity*: Universities should ensure equitable access to licensed CAT tools, adequate lab facilities, and reliable internet to close the digital divide. Special support should be extended to rural learners and under-resourced institutions. Disparities in tool access among universities affect the quality of training (Al-Rumaih, 2021).

(6) *Foster industry-academia collaboration*: Educational institutions should partner with translation companies and AI developers to ensure that translation training reflects real-world market requirements, encouraging collaborative projects and internships. For example, academic training should be aligned with localization and project-based industry workflows (Odacıoğlu & Kokturk, 2015)

(7) *Support longitudinal research and evaluation*: Further research is needed to assess the long-term impact of AI on students' translation quality, critical thinking, and job market success. Mixed-methods studies and longitudinal data will deepen the understanding of AI's pedagogical values. For example, there exist gaps in empirical evidence on AI post-editing's educational effects (Omar & Salih, 2024).

This systematic review provides insights into AI translation tools in tertiary translation education, but some limitations exist. Firstly, it focuses on studies published in English, potentially excluding valuable research from other languages and contexts, especially studies from non-Western educational contexts where AI translation tools may be used differently. Additionally, the varied research designs limit the ability to draw broad conclusions. These limitations highlight the need for diverse, longitudinal, and contextually rich research in future studies. Future research may include studies published in multiple languages and sourced from a wider range of databases to capture various educational contexts, particularly non-Western ones. The employment of standardized evaluation frameworks can improve comparability across varied study designs with mixed-methods and longitudinal approaches. Also, socio-cultural and institutional factors should be considered to provide more contextually grounded and globally relevant insights into the use of AI translation tools in tertiary education.

Based on the findings from this systematic review, we can see that AI translation tools serve as valuable teaching and learning supplements when employed thoughtfully, ethically, and systematically, although they cannot replace human expertise in translation practice. The impact of these tools on translation education depends much on how universities, educators, and students utilize them as instruments of empowerment rather than mere shortcuts to automation.

Acknowledgments

This research was funded by the University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City, as part of a university-level student research project for the 2024–2025 academic year, based on Decision No. 341/TB-XHNV-ĐN&QLKH dated March 31, 2025.

References

Alharbi, W. (2023). The use and abuse of artificial intelligence-enabled machine translation in the EFL classroom: An exploratory study. *Journal of Education and E-Learning Research, 10*(4), 689–701. https://doi.org/10.20448/jeelr.v10i4.5091

Al-Rumaih, L. A. (2021). The integration of computer-aided translation tools in translator-training programs in Saudi universities: Toward a more visible state, *Arab World*

English Journal for Translation & Literary Studies, 5(1), 336-362. https://dx.doi.org/10.24093/awejtls/vol5no1.23

Bakhov, I., Bilous, N., Saiko, M., Isaienko, S., Hurinchuk, S., & Nozhovnik, O. (2024). Beyond the dictionary: Redefining translation education with artificial intelligence-assisted app design and training, *International Journal of Learning, Teaching and Educational Research, 23*(4), 118-140. https://doi.org/10.26803/ijlter.23.4.7

Bao, V. N., & Thi, T. C. N. (2021). Using the internet for self-study to improve translation for English-majored seniors at Van Lang University. *International Journal of TESOL & Education, 1*(2), 110-147. EOI: http://eoi.citefactor.org/10.11250/ijte.01.02.007

Bates, T., Cobo, C., Mariño O., & Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education, 17:*42. https://doi.org/10.1186/s41239-020-00218-x

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

Bouguesmia, M. T. (2020). Using AI in translation: A technological leap, or a translator's nightmare. ALTRALANG Journal, 2(2), 78–102. https://www.asjp.cerist.dz/en/article/137611

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic Approaches to a Successful Literature Review* (2nd ed.). SAGE Publications Ltd. https://www.amazon.com/Systematic-Approaches-Successful-Literature-Review/dp/0857021354

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science, 136*, 16-24. https://doi.org/10.1016/j.procs.2018.08.233

Deng, X., & Yu, Z. (2022). A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability, 14*(13), 7598. https://doi.org/10.3390/su14137598

Gough, D., Oliver, S., & Thomas, J. (2017). *An Introduction to Systematic Reviews*. SAGE Publications Ltd. https://uk.sagepub.com/en-gb/asi/an-introduction-to-systematic-reviews/book245742

Han, B. (2020). Translation, from pen-and-paper to computer-assisted tools (CAT Tools) and machine translation (MT). In *Proceedings of the 14th International Conference INTER-ENG 2020,* 63, 56-62. MDPI. https://www.mdpi.com/2504-3900/63/1/56

Hazaea, A. N., & Qassem, M. (2024). Translation Competence in Translator Training Programs at Saudi Universities: Empirical Study. *Open Education Studies.* 6: 20240029. Retrieved from https://www.degruyter.com/document/doi/10.1515/edu-2024-0029/html

Kenny, D., & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer, 8*(2), 276-294. https://doi.org/10.1080/1750399X.2014.936112

Koka, N. A. (2024). The integration and utilization of artificial intelligence (AI) in supporting older/senior lecturers to adapt to the changing landscape in translation pedagogy. *Migration Letters, 21*(S1), 59-71. https://www.migrationletters.com/index.php/ml/article/view/3153

Laksana, K. N., & Komara, C. (2024). Indonesian EFL students' perceptions of DeepL machine translation tool: Utilization, advantages, and disadvantages. *Journal of Language and Literature Studies, 4*(2), 256-276. https://doi.org/10.36312/jolls.v4i2.1931

Le, T. K. V., & Dao, T. M. T. (2019). An investigation into students' perception of utilizing online dictionaries in translation-interpretation. *Revista ESPACIOS, 40*(39), 2-18. https://www.revistaespacios.com/a19v40n39/in194036.html

Li, X., Gao, Z., & Liao, H. (2024). An empirical investigation of college students' acceptance of translation technologies. *PLOS ONE, 19*(2), e0297297. https://doi.org/10.1371/journal.pone.0297297

Liu, K., & Afzaal, M. (2021). Artificial intelligence and translation teaching: A critical perspective on the transformation of education. *International Journal of Educational Sciences, 33*(1-3), 64-73. https://doi.org/10.31901/24566322.2021/33.1-3.1159

Moher, D., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Odacıoğlu, M. C., & Kokturk, S. (2015). The effects of technology on translation students in academic translation teaching. *Procedia - Social and Behavioral Sciences, 197*, 1085-1094. https://doi.org/10.1016/j.sbspro.2015.07.349

Omar, A., Khafaga, A. F., & Shaalan, I. E. A. W. (2020). The impact of translation software on improving the performance of translation majors: A case study of the Saudi universities. *International Journal of Advanced Computer Science and Applications, 11*(4), 287-292. https://doi.org/10.14569/IJACSA.2020.0110437

Omar, L. I., & Salih, A. A. (2024). Systematic review of English/Arabic machine translation postediting: Implications for AI application in translation research and pedagogy. *Informatics, 11*(2), 23. https://doi.org/10.3390/informatics11020023

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Wiley-Blackwell.

Phuong P. X. H. (2024). Using ChatGPT in English language learning: A study on I.T. students' attitudes, habits, and perceptions. *International Journal of TESOL & Education, 4*(1), 55-68. https://doi.org/10.54855/ijte.24414

Roza, V., & Zulhirawati. (2023). Higher students' perception of using ChatGPT in translating

English texts. *Bukittinggi International Conference on Education (BiCED) Proceedings, 1*, 65-73. https://doi.org/10.30983/biced.v1i1.278

Soysal, F. (2023). Enhancing translation studies with artificial intelligence (AI): Challenges, opportunities, and proposals. *International Journal of Philology and Translation Studies, 5*(2), 178-191. https://doi.org/10.55036/ufced.1402649

Tavares, C., Tallone, L., Oliveira, L., & Ribeiro, S. (2023). The challenges of teaching and assessing technical translation in an era of neural machine translation. *Education Sciences, 13*(6), 541-559. https://doi.org/10.3390/educsci13060541

Tian, Xujun. (2024). Personalized translator training in the era of digital intelligence: Opportunities, challenges, and prospects. *Heliyon, 10*, e39354. https://doi.org/10.1016/j.heliyon.2024.e39354

Yanti, M., & Meka, L. M. C. (2019). The students' perception in using Google Translate as a media in translation class. In *Proceedings of International Conference on English Language Teaching (INACELT)*, *3*(1), 128-146. https://doi.org/10.32528/issh.v2i1.120

Zhang, J. (2023). Exploring undergraduate translation students' perceptions towards machine translation: A qualitative questionnaire survey. In *Proceedings of Machine Translation Summit XIX*, *2: Users Track*, 1-10. Macau SAR, China. Asia-Pacific Association for Machine Translation. https://aclanthology.org/2023.mtsummit-users.1/

Biodata

Nguyen Thi Nhu Ngoc, Ph.D. is currently Vice-dean cum Chair of the Department of Translation and Interpreting at the Faculty of English Linguistics and Literature, USSH, VNUHCM. She has been an English teacher and a part-time translator since 1997. Her main research interests are English Teaching, Translation Studies, Comparative Linguistics, and Intercultural Communication.

Tran Thanh Truc is currently an undergraduate student majoring in Translation and Interpreting at the Faculty of English Linguistics and Literature, USSH, VNUHCM. Her academic interests include Translation Studies, Interpreting, and Literature.

Nguyen Ngoc Hoang An is currently an undergraduate student majoring in Translation and Interpreting at the Faculty of English Linguistics and Literature, USSH, VNUHCM. Her academic interests include Translation Studies, Sociolinguistics, and Literature.

Lam Hoang Phat is currently an undergraduate student majoring in British-American Culture & Literature at the Faculty of English Linguistics and Literature, USSH, VNUHCM. His academic interests include Intercultural Communication, British-American Identity & Lifestyle, British-American Poems, and Short Stories.

Nguyen Hua My San is currently an undergraduate student majoring in the Department of English Linguistics and Language Teaching at the Faculty of English Linguistics and Literature, USSH, VNUHCM. Her academic interests include English Teaching, Linguistics, and Intercultural Communication.

Tran Nguyen Anh Thu is currently an undergraduate student majoring in Translation and Interpreting at the Faculty of English Linguistics and Literature, USSH, VNUHCM. She also serves as President of the Faculty's Student Association. Her academic interests include Translation Studies, Sociolinguistics, and Literature.

Factors Affecting the Outcomes of a Teacher Training Program in Competency-Based Student Assessment In Vietnam

Nguyen Thi Hong Nhat[1*], Le Ha Thanh[2], Bui Minh Duc[3]

[1]Faculty of English, Hanoi Pedagogical University 2, Vietnam
[2]Faculty of Psychology and Pedagogy, Hanoi Pedagogical University 2, Vietnam
[3]Faculty of Philology, Hanoi Pedagogical University 2, Vietnam
[*]Corresponding author's email: nguyenthihongnhat@hpu2.edu.vn
[*]https://orcid.org/0000-0003-2894-3687
https://doi.org/10.54855/ijte.25528

Received: 13/01/2025 Revision: 18/05/2025 Accepted: 22/05/2025 Online: 23/05/2025

ABSTRACT

The study was conducted to examine the factors influencing the outcomes of a training programme designed for secondary and high school teachers in Northern midland and mountainous provinces. Based on the analysis, the researchers proposed recommendations for enhancing the effectiveness of teacher professional development programmes. Data were collected through a questionnaire with 1422 responses from core secondary and high school teachers in seven northern provinces of Vietnam (Vinh Phuc, Phu Tho, Tuyen Quang, Ha Giang, Yen Bai, Lao Cai, and Bac Giang). These teachers have completed a training module on assessment and testing. It was obvious from the investigation that male teachers from non-ethnic majority groups outperformed female teachers and teachers from ethnic minority communities in terms of the results they achieved for the training practices. Furthermore, the linear regression analysis indicated that the following factors: organizing of training, assessment methods, training materials and training contents, affected the training quality from the highest to the lowest, respectively. The paper concludes with some recommendations to address the constraints and improve the quality of training programmes for high school teachers so that they can meet the needs of education innovation in the coming years.

Keywords: Affecting factors, training outcomes, assessment, Competency-based education, Vietnam

Introduction

Competency-based education is the new idea of the General Education Program (GEP) 2018 in Vietnam. This is the core ideology throughout the program, which results in many significant changes to educational activities in general and teaching, testing and assessment in particular. To help teachers effectively implement the GEP 2018, the Ministry of Education and Training,

CITATION | Nguyen, T. H. N., Le, H. T., & Bui, M. D. (2025). Factors Affecting the Outcomes of a Teacher Training Program in Competency-Based Student Assessment In Vietnam. *International Journal of TESOL & Education*, *5*(2), 149-162. DOI: https://doi.org/10.54855/ijte.25528

through the ETEP Program, has implemented Module 3: Competency-based assessment of primary/secondary/high school students (MOET, 2020). This module also aims to strengthen teachers' long-term and sustainable professional capacity, meeting modern education's requirements in the context of integration and internationalization. The cooperation between the key teacher education universities and the Departments of Education and Training, primary and secondary schools nationwide has implemented it. Hanoi Pedagogical University 2 (HPU2) is one of the seven pedagogical universities assigned by the Ministry of Education and Training to foster teachers under the ETEP Program for 07 northern midland and mountainous provinces: Vinh Phuc, Phu Tho, Tuyen Quang, Ha Giang, Yen Bai, Lao Cai, and Bac Giang.

The practical experience in teacher training in the last three years of implementing GEP 2018 and research-based evidence on the outcomes of teacher training have scientific significance for teacher training institutions and the entire educational sector in Vietnam. This was especially important since GEP 2018 continues to be implemented in the coming years. Therefore, conducting preliminary studies to identify the factors influencing the outcomes of teacher training, including training on competency-based assessment, is essential for the better implementation of GEP 2018 in the upcoming years.

Literature Review

Different factors influencing the outcomes of teacher training have been identified in several studies worldwide and in Vietnam. In a comprehensive study on influencing factors, Prenger, Poortman, and Handelzalts (2017) synthesized various key factors including: teachers' career development motivation; convergence of development goals and leadership; the content of training that is closely linked to students' outcomes; knowledge-sharing support within the school environment; and support from stakeholders (principals and colleagues). Voogt et al. (2011) further emphasized that both the quantity and quality of support from stakeholders are crucial for achieving meaningful changes. Stoll et al. (2006) pointed out that the lack of support from the school hindered the application of the trained content into the teachers' work. Another study conducted in Ghana revealed that teachers primarily engaged in training through workshops, distance learning, and regular training sessions. However, these forms of training rarely met the needs of the teachers. Therefore, a broader policy framework is needed to guide teacher training provision, participation, and implementation (Abakah et al., 2022). In the context of Vietnam, Nguyen (2022) found that teacher agency is critical in influencing training program effectiveness. Nguyen and Truong (2021) added that "relationships" and "environments and structures" were particularly influential factors on training effectiveness.

Related to the challenges influencing teachers' engagement in professional development, Geldenhuys and Oosthuizen (2015) conducted research in the context of South Africa and identified four issues of concern: inadequate contribution of school management to teacher training; teachers' hesitance to participate in training activities; and insufficient training content in training programs. The study recommended the involvement of teachers in planning, implementing, and evaluating training programs to enhance their effectiveness. Other studies have also added that the success of training courses comes from teachers' autonomy (Nguyen, 2019), a supportive mechanism and a collaborative school culture for successful

experimentation, support from school leadership (Nicolaidis & Mattheoudakis, 2008), respect for and assessment of the professional expertise of participating teachers (Pyle, et al., 2011), a sense of safety to try new things (Guskey, 2000), and the availability of appropriate resources (Nguyen, 2018; Seçer, 2010).

Regarding the impact of assessment-focused professional development on teachers, numerous studies have been conducted with various subjects and scales (Andersson & Palm, 2017; Randel et al., 2016). For instance, Andersson and Palm (2017) conducted a study with 22 4th-grade Math teachers in Sweden to investigate the impact of a professional development program on teachers' changes in assessment practices and their influence on students' achievements. The results showed that teachers underwent significant changes after participating in the training, leading to improved learning outcomes for students in their classes. In another study, Randel et al. (2016) conducted a research on the impact of a professional development program focused on formative assessment implementation in classrooms. The researchers randomly assigned 67 elementary schools to receive either formative assessment implementation training or continue regular professional development. Teachers in these schools formed study groups and were provided with instructional materials. They were suggested to implement formative assessments in their classrooms under real conditions, without the researchers' direct involvement or requirements. Analyzing all schools and 9,596 students, no significant impact of the training was found on students' math scores in statewide assessments. Impact analyses with 231 teachers revealed positive effects of the training on teachers' assessment knowledge and the frequency of student participation in classroom assessments. However, the training did not clearly impact teachers' assessment practices.

In a recent study, Schelling and Rubenstein (2023) evaluated the outcomes of teacher training in assessment for elementary school teachers. Among the surveyed teachers (n = 283), only 56% of teachers had received assessment training during their university courses, whereas 84% had received training through specialized professional development courses. Quantitative results indicated that the frequency of assessment training was positively related to teachers' competency, attitude, and practices in assessment. It can be observed that these studies highlight the significance of assessment-focused training and professional development for teachers.

In Vietnam, research on teacher training in response to GEP 2018 focuses primarily on theoretical issues. Bui (2018) advocates for redesigning the training process, prioritizing educational research, and extending practical training periods for professional development. Identifying training needs and selecting appropriate models are critical for effective teacher development. Pham and Nguyen (2016) recommend improving capabilities of higher education faculty and secondary school teachers, upgrading infrastructure, establishing collaboration between institutions and schools, developing policy frameworks, and creating training resources. Nguyen (2022) identifies that student factors and access conditions influence online teaching effectiveness in Hanoi high schools. This research proposes five solutions to enhance online teaching based on both positive and negative factors affecting implementation.

A general overview of other studies in Vietnam regarding the factors influencing teachers' professional activities shows that the main focus of research is on surveys, analyses, and assessments of the current situation concerning factors affecting teachers' teaching and

assessment activities, particularly about specific subjects and educational contexts. The overview has illustrated the research landscape regarding teacher professional development in general and the factors influencing the outcomes of teacher training in particular. This study further clarifies the factors influencing the outcomes of teacher training concerning competency-based student assessment in Vietnam through a specific training program - the ETEP program - in the provinces of the northern midland and mountainous region.

Research Questions

From the context mentioned above, this study focuses on answering the following research questions:

1. What are the factors affecting the training effectiveness of a teacher training program in competency-based student assessment?
2. What is the degree of influence of each factor?

Methods

Participants

Table 1 shows the characteristics of the participants in this study. The participants were a convenience sample recruited from 7 provinces: Ha Giang, Bac Giang, Lao Cai, Vinh Phuc, Yen Bai, Phu Tho, and Tuyen Quang.

Table 1

Characteristics of participants

Characteristics		**N**	**%**
Sex	Male	640	45
	Female	782	55
Difficult area	No	847	59.6
	Yes	575	40.4
Ethnic minorities	No	1274	89.6
	Yes	148	10.4
Teacher level	Secondary school	1209	85
	High school	213	15
Do management	No	1352	95.1
	Yes	70	4.9
Years of work	< 15 years	747	52.5
	≥ 15 years	675	47.5
School location	Tuyen Quang	167	11.7
	Ha Giang	206	14.5
	Lao Cai	202	14.2
	Yen Bai	181	12.7
	Phu Tho	268	18.8
	Bac Giang	237	16.7
	Vinh Phuc	161	11.3

A total of 1,422 teachers took part in the study. The majority of participants were female, comprising 55% (n=782) of the sample. Additionally, the majority of teachers were not from challenging areas, accounting for 59.6% (n=847) of the total. Among the participants, 85% (n=1,209) were secondary school teachers, and 52.5% (n=747) had less than 15 years of teaching experience. Within the participant pool, 10.4% (n=148) identified as ethnic minorities, while 4.9% (n=70) were responsible for managerial roles.

Survey method

The survey was conducted during the stage of organizing the training of Module 3 for core high school teachers in the ETEP program, in an online form. The survey consists of two parts: Part one collected demographic information of participants such as gender (Male/Female), difficult area (No/Yes), ethnic minorities (No/Yes), teacher level (Secondary/High school), teacher do managerment (No/Yes), years of work (<15 years/≥ 15 years), and school location. Part two collected teachers' self-reports after participating in the training on "Training effectiveness", "Training materials", "Software and IT for training", "Training organization", and "Online support after training".

Survey instruments

Teachers evaluated "Training effectiveness " by 6 questions: (1) "Achieve personal needs for professional development", (2) "Help me improve to achieve Teacher's Professional Standards" , (3) "I am confident enough to support my colleagues with the contents of the training", (4) "I am confident in my ability to support colleagues in the activities related to the implementation content to the module", (5) "I feel I am a member of the learning community built up by the training course" and (6) "I am satisfied with the quality of the training module". The questions are graded on a Likert scale, from 1 = "Strongly disagree" to 4 = "Strongly agree". Cronbach's Alpha for the "Training effectiveness " scale was 0.84. The mean score of the scale was divided into 4 levels, from completely disagree (1 point) to strongly agree (4 points):

Level 1 (Very low): 1 to 1.75

Level 2 (Low): 1.75 to 2.5

Level 3 (Pass): 2.5 to 3.25

Level 4 (High): 3.25 to 4

The higher the mean score, the better the teacher's agreement about the training effectiveness. In this study, the mean scale "Training effectiveness " score was 3.30 (SD: 0.46) at level 4.

Factors affecting the training effectiveness

In this study, we evaluated 4 factors that affect the training results (Table 2). We use a 4-point Likert scale from 1 = "Strongly disagree" to 4 = "Strongly agree".

Table 2

Reliability and characteristics of the scales

Scales	Cronbach's Alpha	No of items	Mean	SD
(1) Training materials	0.83	4	3.32	0.48
(2) Software and IT for training	0.85	6	3.31	0.47
(3) Training organization	0.84	7	3.28	0.46
(4) Online support after training	0.86	5	3.31	0.48

The analysis results of 4 scales showed that the Cronbach's Alpha reached the hight level (>0,8), proving that they were suitable for data analysis.

Data analysis

Baseline demographics were analyzed for descriptive purposes. Group comparisons of the "Training effectiveness " were made using t-test and Chi-square tests. Correlation analysis was performed to consider the relationship and influence of the variables "Training materials", "Software, IT for training", "Training organization", and "Online support after training" on "Training effectiveness ". The coefficient β (Beta) was calculated at the significance level of 0.05.

Results

Teachers' self-reported on training effectiveness

The study aimed to assess the effectiveness of Module 3 training by examining self-reported data from teachers who had completed the program. Overall, participants rated the training effectiveness at level 4, with a mean of 3.30 and a standard deviation of 0.46. To delve into variations among teacher characteristics, we conducted t-tests and one-way ANOVA tests, with statistically significant differences indicated by p-values in bold.

Examining the results by gender, male teachers rated the training more positively than their female counterparts (Mean: 3.36 vs 3.25, $p < 0.05$). Similarly, teachers who were not from ethnic minorities perceived the training as more effective compared to those who identified as ethnic minorities (Mean: 3.31 vs 3.20, $p < 0.01$). High school teachers also rated the training higher than secondary school teachers (Mean: 3.37 vs 3.29, $p < 0.05$). The training effectiveness of Module 3 for secondary school teachers was shown in Table 3.

Table 3

Training effectiveness

Variables		Training effectiveness	
		Mean	**p**
Sex[a]	Male	3.36	**0.000**
	Female	3.25	
Difficult area[a]	No	3.32	0.151
	Yes	3.28	
Ethnic minorities[a]	No	3.31	**0.008**
	Yes	3.20	
Teacher level[a]	Secondary school	3.29	**0.016**
	High school	3.37	
Do management[a]	No	3.30	0.511
	Yes	3.33	
Years of work[a]	< 15 years	3.29	0.631
	≥ 15 years	3.30	
School location[b]	Tuyen Quang	3.32	**0.003**
	Ha Giang	3.31	
	Lao Cai	3.29	
	Yen Bai	3.17	
	Phu Tho	3.34	
	Bac Giang	3.28	
	Vinh Phuc	3.37	

Note:
[a] t-test;
[b] One way ANOVA test;
P-values in bold indicate a statistically significant difference

However, no statistically significant differences were found in the assessment of training effectiveness based on difficulty areas (No/Yes), management roles (No/Yes), and years of work (<15 years/≥ 15 years). This suggests that the training program's impact was consistent across teachers working in diverse difficulty areas, holding management roles, and varying levels of teaching experience.

The study findings highlight variations in teachers' perceptions of the effectiveness of Module 3 training based on gender, ethnicity, and school level. This finding is consistent with that of Prenger, Cindy, and Adam (2017), identified various key factors influencing teachers' perceptions of training quality. These insights can inform future training initiatives to tailor approaches and content to better meet the diverse needs of teachers in different contexts.

Factors affecting the training effectiveness

Four factors affecting the training results of Module 3 were considered: (1) "Training materials", (2) "Software and IT for training", (3) "Organization of training", and (4) "Online support after training". First, it is necessary to consider the correlation between these factors

with the training effectiveness. Table 4 shows the correlation between the four factors and the training effectiveness.

Table 4

Correlation between training effectiveness and influencing factors [a]

	(1)	(2)	(3)	(4)	(5)
(1) Training effectiveness	1				
(2) Training materials	0.83**	1			
(3) Software and IT for training	0.84**	0.84**	1		
(4) Training organization	0.85**	0.85**	0.85**	1	
(5) Online support after training	0.89**	0.89**	0.84**	0.85**	1

Note:
** $p<0.01$
[a] Spearman correlation coefficient

Four factors are highly positively correlated with statistical significance with training effectiveness ($r>0.8$; $p<0.01$). To confirm the influence of these factors on the training effectiveness, we carried out a linear regression analysis. Dependent variable is "training effectiveness", and independent variables are (1) “Training materials”, (2) “Software and IT for training”, (3) “Organization of training”, and (4) “Online support after training”. The results of the linear regression are shown in Table 5.

Table 5

Results of linear regression analysis[b]

Model	R	R^2	R^2_{adj}	Standard error	Durbin-Watson
1	0.934[a]	0.873	0.873	0.166	1.949

Note:
[a] Predictive factors: Training materials, Software and IT for training, Organization of training, Online support after training.
[b] Dependent variable: Training effectiveness

The R-value of 0.93 shows that the relationship between the variables in the model has a high correlation. The regression results report of the model shows that the value of R2 = 0.87, which means that the model's relevance is 87.3%, or in other words, 87.3% of the variation of the model. The variable "training effectiveness" is explained by four components. The R^2_{adj} value is 0.87, which means a linear regression model between training effectiveness and four factors. Durbin-Watson coefficient (DW) confirmed that the model does not violate when using the multiple regression method because the obtained DW value is 1.95 (ranges from 1 to 3). Thus, the multiple regression model meets the evaluation and appropriateness test requirements for drawing research results.

Table 6

Regression coefficients in the model

Independent variables (X)	Dependent variables (Y)	Beta (β)	t	p
Constant	Training effectiveness		4.45	0.000
Training materials		0.098	4.34	0.000
Software and IT for training		0.136	5.41	0.000
Organization of training		0.212	8.59	0.000
Online support after training		0.525	21.76	0.000

The standardized regression values of the independent variables in the model: Training materials is 0.098; Software and IT for training is 0.136; Training organization is 0.212; and Online Support after training is 0.525.

Regression equation for the model:

$$Y = 0.098X_1 + 0.136\ X_2 + 0.212\ X_3 + 0.525\ X_4$$

The model explains 87.3% of the variation of the variable Y, the remaining 22.50% of the variation is explained by other variables outside the model. The model shows that the independent variables positively influence the training effectiveness at the 99% reliability level. The regression equation shows that the evaluation score of "Training materials" increases to 1, the training effectiveness increases by 0.098. Similarly, when the evaluation score of "IT software for training" increases by 1 point, the training efficiency increases by 0.136; when "Organization of training" increases by 1, the training effectiveness increases by 0.212; when "Online support after training" increases by 1, the training effectiveness increases by 0.525.

The above analysis concludes that the theoretical model is suitable for the research data and four factors affecting the training effectiveness of Module 3. The results of testing the theoretical model are shown in Figure 1.

Figure 1.

Factors affecting the effectiveness of Training Module 3

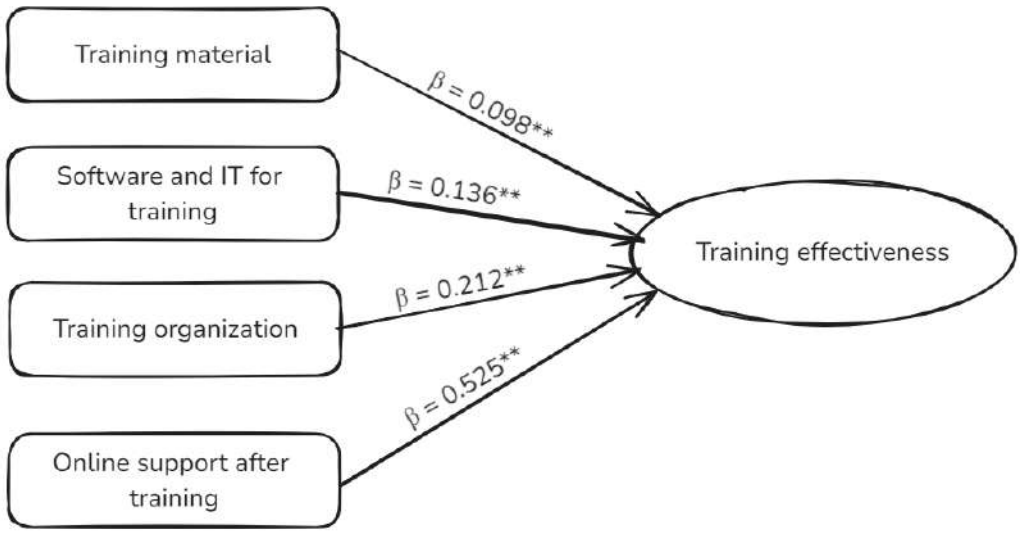


Figure 1 shows that the training effectiveness was most influenced by the "Online support after the training" factor (Beta = 0.525, $p<0.001$); the second is the "Organization of training" (Beta = 0.212, $p<0.001$); the third is "Software and IT for training" (Beta = 0.136, $p<0.001$) and the fourth is "Training materials" (Beta = 0.098, $p<0.001$).

Therefore, the results of regression model testing show that all four independent variables have a positive influence on training effectiveness. This study supports evidence from previous observations (e.g., Nguyen, 2019; Pyle et al., 2011) that teacher training and retraining institutions need to make efforts to improve these factors to improve the effectiveness of teacher training.

Discussion

The findings of this study offer valuable insights into the factors influencing the outcomes of teacher training programs in competency-based student assessment within the Vietnamese context. Consistent with prior literature, the results affirm that individual and contextual factors significantly shape teachers' perceptions of training effectiveness. More specifically, the study identified a significant disparity in training perception between ethnic majority and minority teachers, with the latter group rating the program less favorably. This aligns with Nguyen's (2022) conclusion that individual and contextual access issues—particularly in underserved regions—can hinder teachers' engagement and learning outcomes in professional development. It also suggests potential language or cultural barriers that may not have been adequately addressed in the training content or delivery. Teachers' professional level (secondary vs. high school) emerged as another influential factor. High school teachers rated the training higher than secondary school teachers, suggesting that the content or pedagogical orientation of Module 3 may have been better aligned with high school curricula or expectations. This calls attention to the importance of aligning training content with the specific instructional contexts of different school levels, as emphasized by Voogt et al. (2011) and Abakah et al. (2022). Interestingly, variables such as working in difficult areas, having management roles, and years of teaching experience did not produce significant differences in the perceived effectiveness of the training. This is a noteworthy finding as it suggests that the ETEP training may have achieved a degree of consistency in delivery across various professional and regional contexts. However, the lack of significant differentiation also raises questions about whether the program was sufficiently responsive to the specific challenges faced by teachers in particularly disadvantaged areas—an issue highlighted by Stoll et al. (2006) and Pham & Nguyen (2016).

The findings support the idea that individual characteristics and school-level contexts, rather than administrative roles or seniority, play a more central role in shaping the outcomes of teacher training in competency-based assessment. This resonates with the findings of Geldenhuys and Oosthuizen (2015), who stressed the importance of teacher involvement and contextual alignment in the planning and implementation of training.

The results emphasize the importance of contextual responsiveness, cultural sensitivity, and alignment of content with teaching levels in teacher training programs. These findings reinforce the call from Vietnamese scholars such as Bui (2018) and Nguyen (2022) for more needs-based

and regionally adaptable training models. To enhance the overall effectiveness and equity of training initiatives, future efforts should consider differentiated strategies that take into account the diverse backgrounds and needs of teachers, particularly in ethnically diverse and resource-constrained settings.

The analysis also indicates that the theoretical model is suitable for the research data, and all four factors significantly and positively affect the training effectiveness of Module 3. The results strongly suggest that "Online support after training" is the most influential factor among those examined, followed by "Organization of training".

These findings align with and support evidence from previous observations (e.g., Nguyen, 2019; Pyle et al., 2011), which emphasize the importance of various support mechanisms and organizational aspects in teacher training. The study underscores the need for teacher training institutions to focus efforts on improving these specific factors, particularly post-training support and organization aspects, to enhance the overall effectiveness of teacher professional development programs like the one studied.

Conclusion

In conclusion, this study explored the determinants influencing teacher training programs' effectiveness, specifically focusing on the development of students' quality and capacity. The empirical testing of our theoretical model revealed four key factors that significantly impact training effectiveness: Online support post-training, Training organization, Software and IT infrastructure for training, and Training materials.

Moreover, the research illuminated noteworthy disparities in the evaluation of training effectiveness across different teacher demographics. Male teachers, non-ethnic minority educators, and those in high schools reported higher levels of satisfaction with the training compared to their female counterparts, ethnic minority teachers, and secondary school instructors. While the study provides insights into the differences in evaluating training effectiveness among different teacher demographics, it has some limitations. First, our data was collected using the self-reported method; therefore, the responses may be biased due to social desirability. Second, this study was performed on data collected at a single point in time, so caution should be exercised when interpreting the results of analyses based on cross-sectional data.

In response to these findings, we put forth several recommendations for educational institutions. Emphasizing the enhancement of online support post-training and optimizing the organization of training programs are crucial steps. Investing in upgrading software and information technology infrastructure for training is necessary to facilitate a more efficient and modernized learning experience. Additionally, improving the quality of training materials is imperative, aligning them with practical and effective strategies that conform to the evolving landscape of general education innovation as mandated by the 2018 education program.

Similarly, for departments of education and training, high schools, and secondary schools, the study suggests tailoring training to the specific needs of teaching staff, especially among female

teachers and those from ethnic minorities. Collaborating with teacher education universities for regular exchange and coordination in organizing online training and post-teacher training support is paramount. This collaboration should also focus on enhancing the information technology infrastructure to ensure a more convenient learning environment. Lastly, involving qualified teachers in the active creation of training materials is recommended, ensuring they are well-suited to practical applications and align with the requirements of the 2018 education program.

By implementing these recommendations, educational institutions and relevant departments can contribute to continuously improving teacher training programs, meeting the dynamic needs of education innovation in the years to come.

Acknowledgments

The authors would like to thank all colleagues and school teachers who participated in this study for their support.

References

Abakah, E., Widin, J., & Ameyaw, E. (2022). Continuing Professional Development (CPD) Practices Among Basic School Teachers in the Central Region of Ghana. *SAGE Open, 12*, 215824402210945. https://doi.org/10.1177/21582440221094597

Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction, 49*, 92-102. doi: https://doi.org/10.1016/j.learninstruc.2016.12.006

Bui, V. Q. (2018). *Teacher training to meet the requirements of the new general education curriculum: Some theoretical and practical issues [Bồi dưỡng giáo viên đáp ứng yêu cầu chương trình giáo dục phổ thông mới – một số vấn đề lí luận và thực tiễn]*. Paper presented at the National Scientific Conference Proceedings on Innovation in Teacher Education and Training to Meet the New General Education Curriculum, Professional Standards, and Local Labor Market Demands, Hanoi.

Geldenhuys, J. L., & Oosthuizen, L. C. (2015). Challenges influencing teachers' involvement in continuous professional development: A South African perspective. *Teaching and Teacher Education, 51*, 203-212. doi: https://doi.org/10.1016/j.tate.2015.06.010

Guskey, T. R. (2000). *Evaluating professional development*: Corwin press.

Ministry of Education and Training (MOET). (2020). *Training manual Module 3: Assessment of upper secondary students based on competency and quality development*. https://www.studocu.vn/vn/document/truong-dai-hoc-vinh/tu-dong-hoa/tai-lieu-mo-dun-3-mon-dia-li-thpt/95517640

Nguyen, D. N. (2022). Commitment to Core Values in Professional Development in ELT Contexts: A Case Study. *International Journal of TESOL & Education*, *2*(5), 100-109. https://doi.org/10.54855/ijte.22257

Nguyen, K. N. A. ., & Truong, D. K. (2021). Exploring the Impacts of Doing Action Research on EFL Teachers' Professional Identities from an Ecological Perspective. *International Journal of TESOL & Education*, *1*(3), 65-87. https://i-jte.org/index.php/journal/article/view/87

Nguyen, T.H.N. (2018). Vietnamese university teachers' views on a large-scale professional development course on using Computer-Assisted Language Learning. In K. Hashimoto & T. Nguyen (Eds.), Professional development of English language teachers in Asia: Lessons from Japan and Vietnam (pp. 130-148). New York, NY: Routledge

Nguyen, T. H. N. (2019). Teachers' implementation of computer-assisted language learning in the context of educational change in Vietnam. In H. T. Nguyen & H. Q. Le (Eds.), *Building teacher capacity in English language teaching in Vietnam* (pp. 133–149). Routledge.

Nguyen, T.T.T. . (2022). Một số yếu tố ảnh hưởng đến dạy học trực tuyến của giáo viên trung học phổ thông tại Hà Nội: Thực trạng và giải pháp. *Tạp chí Khoa học Giáo dục, Viện Khoa học Giáo dục Việt Nam, 2*, 53-58.

Nicolaidis, K., & Mattheoudakis, M. (2008). Utopia vs. reality: The effectiveness of in-service training courses for EFL teachers. *European Journal of Teacher Education - EUR J TEACH EDUC, 31*, 279-292. https://doi.org/10.1080/02619760802208460

Pham, H.Q. , & Nguyen, D.N. . (2016). Nâng cao hiệu quả bồi dưỡng giáo viên đáp ứng yêu cầu đổi mới căn bản và toàn diện giáo dục. *Tạp chí Giáo dục, số đặc biệt tháng 10/2016*, 2-5.

Prenger, R., Poortman, C., & Handelzalts, A. (2017). Factors influencing teachers' professional development in networked professional learning communities. *Teaching and Teacher Education, 68*, 77-90. https://doi.org/10.1016/j.tate.2017.08.014

Pyle, A., Wade-Woolley, L., & Hutchinson, N. (2011). "Just listen to us": The role of teacher empowerment in the implementation of responsiveness to intervention. *Alberta Journal of Educational Research, 57*, 258-272.

Randel, B., Apthorp, H., Beesley, A., Clark, T., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research, 109*, 1-12. https://doi.org/10.1080/00220671.2014.992581

Schelling, N., & Rubenstein, L. (2023). Pre-service and in-service assessment training: Impacts on elementary teachers' self-efficacy, attitudes, and data-driven decision making practice. *Assessment in Education Principles Policy and Practice*. https://doi.org/10.1080/0969594X.2023.2202836

Seçer, Z. (2010). An analysis of the effects of in-service teacher training on Turkish preschool teachers' attitudes towards inclusion. *International Journal of Early Years Education, 18*, 43-53. https://doi.org/10.1080/09669761003693959

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional Learning Communities: A Review of the Literature. *Journal of Educational Change, 7*, 221-258. https://doi.org/10.1007/s10833-006-0001-8

Voogt, J., Westbroek, H. B., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., & Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education - Teach Teach Educ, 27*, 1235-1244. https://doi.org/10.1016/j.tate.2011.07.003

Biodata

Dr. Nguyen Thi Hong Nhat is the Dean of the Faculty of English, a teacher educator and a senior lecturer of English at Hanoi Pedagogical University 2, Vietnam. She holds a Doctor degree from the University of Queensland. She has received several scholarships from Vietnamese, Australian and the US government scholarships to complete her higher education study and other professional development courses. She has published research articles in both local and international journals, one book and 5 locally-used course-books. She has been a peer reviewer for some international journals and conferences such as AsiaCALL Online Journal 2021, RELO Vietnam Fullbright TEA Alumni Conference, LLCE 2020 Conference, Vietnam Education Symposium. She has also actively contributed to VietTESOL's professional development for K-12 to organize and deliver PD trainings for many English teachers in Vietnam. Her research interest is in the teaching of listening as a foreign language skill, EFL materials development, textbooks, curriculum development, teacher education, Computer-Assisted Language Learning and educational change.

Dr. Ha Thanh Le graduated with the PhD in Psychology from Hanoi National University (VNU), Hanoi, Vietnam. He has 20 years of teaching and scientific research experience at Hanoi Pedagogical University 2 (HPU2). He is currently working at the Faculty of Psychology and Pedagogy, HPU2, Vinh Phuc, Vietnam. His main research interests are school psychology, mental health in educational institutions, and issues related to student development in the educational institution environment. He can be contacted at email: haltpsy@gmail.com

Assoc. Prof. Dr. Bui Minh Duc is the Dean the Faculty of Philology at Hanoi Pedagogical University 2 in Vietnam. He holds a doctorate from the Hanoi University of Education and received an associate professor degree. He has nearly 25 years of experience in educational research and teacher training. He has participated in many domestic and international projects on teacher education. His research focuses on teacher education, school curriculum development, education management and teaching literature and Vietnamese.

International Journal of TESOL & Education

ISSN: 2768-4563

Vol. 5, No. 2 (2025): TESOL & Education
Doi: https://doi.org/10.54855/ijte.2552
Orcid: https://orcid.org/0000-0002-3233-457X

Publisher: ICTE Press

Address: International Association of TESOL & Education
5243 Birch Falls Ln, Sugar Land, Texas, USA,

Aims and scopes

Aims and scopes

The mission of the IJTE (ISSN: 2768-4563) is to contribute to the literature of instruction by publishing high-quality studies in the areas of TESOL, Education, Computer-Assisted Language Learning (CALL), m-learning, e-learning, curriculum development, learning environments, Pedagogies & Language Teaching, translation and interpretation, teacher education, educational technology, educational developments, quality assurance in education, cultural studies, and other learning-instruction issues. As a peer-reviewed journal, one of our priorities is to ensure high standards of publishing. Publication ethics is an important concern, and editors, peer-reviewers, and authors should pay great attention to meeting ethical standards.

Open Access Policy

IJTE aims to increase visibility and make easier use of open access scientific papers. Readers and their institutions are supported for online access without asking for any royalty, personal information, or log-in details. According to the open-access policy of our journal, all readers are permitted to read, download, copy, distribute, print, link to, and search our articles with no charge.

Copyright

Ethics and Malpractice

The Statement of Publication Ethics & Malpractice

The International Journal of TESOL & Education (ijte) (ISSN: 2768-4563) is committed to upholding ethical standards, retracting and correcting errors. The editorial team's primary responsibility is to discourage publishing malpractice. Any type of unethical conduct is unacceptable, and this Journal's Editorial Team does not tolerate plagiarism in any form. All manuscripts must be the authors' original work and free of indications of plagiarism.

Conflicts of Interests

Authors are requested to disclose interests that are directly or indirectly related to the work submitted for publication. Interests within the last 3 years of beginning the work (conducting the research and preparing the work for submission) should be reported. Interests outside the 3-year time frame must be disclosed if they could reasonably be perceived as influencing the submitted work. Disclosure of interests provides a complete and transparent process and helps readers form their own judgments of potential bias. This is not meant to imply that a financial relationship with an organization that sponsored the research or compensation received for consultancy work is inappropriate. Please download this form.

Disclosures and declarations

All authors are requested to include information regarding sources of funding, financial or non-financial interests, study-specific approval by the appropriate ethics committee for research involving humans and/or animals, informed consent if the research involved human participants, publication on vulnerable populations, and a statement on the welfare of animals if the research involved animals (as appropriate).

The decision on whether such information should be included is not only dependent on the scope of the journal, but also the scope of the article. Work submitted for publication may have implications for public health or general welfare and in those cases, it is the responsibility of all authors to include the appropriate disclosures and declarations.

Compliance with Ethical Standards

To ensure objectivity and transparency in research and to ensure that accepted principles of ethical and professional conduct have been followed, authors should include information regarding sources of funding, potential conflicts of interest (financial or non-financial), informed consent if the research involved human participants, publication on vulnerable populations, and a statement on the welfare of animals if the research involving animals.

Authors should include the following statements (if applicable) in a separate section entitled “Compliance with Ethical Standards” when submitting a paper.

Originality and Plagiarism

Authors can guarantee that their writings are entirely unique and that any writing or words credited to another are properly referenced. Additionally, publications that had an impact on the nature of the finding reported in the manuscript should be acknowledged. The writers must certify that the manuscript has never been written before.

It is not permitted to copy text from other publications without correctly attributing the source (plagiarism) or to publish several papers with nearly identical material by the same writers (self-plagiarism). It is against the law to concurrently apply the same results to more than one Journal. It is prohibited for authors to present results collected from others as their own. Authors should consider the work of those who shaped the direction and course of their research.

Check plagiarism

The ijte will conduct a plagiarism check on all submitted papers using Turnitin’s program. The ijte only considers papers with less than 20% similarity index for publication. IJTE will immediately reject any paper that involves plagiarism.

AI and AI-Assisted Tools

Authorship implies human-only responsibilities and duties. Each author is responsible for ensuring that any questions regarding the accuracy or integrity of any portion of the work are adequately investigated and resolved, and authorship requires the ability to endorse the final version of the work and consent to its submission. The authors are also responsible for ensuring that the work is original, that the indicated authors qualify for authorship, and that the work does not violate the rights of third parties.

Papers generated by AI or AI-assisted Tools are not accepted for publication in this Journal. This policy pertains only to the writing process and not to the use of AI tools to analyze data and gain insights for the research procedure. This policy does not prohibit the use of artificial intelligence and AI-assisted instruments in formal research design or research methods. When AI or AI-assisted tools are used in this context, they should be described in the Methods section as part of the work's methodology.

The authors are responsible and answerable for the content of their work. The authors should indicate in their manuscript if they utilized AI and AI-assisted technologies, and a corresponding statement will appear in the final product. Disclosure of the use of these technologies promotes honesty and trust among authors, readers, reviewers, editors, and contributors and facilitates adherence to the tool's or technology's terms of service.

Brief Information

Country of Publication: Texas, United States
Publisher: ICTE Press, https://i-cte.org/press
Website: https://i-jte.org
Youtube: https://www.youtube.com/@ictepress
Open Access: Yes
ISSN: 2768-4563
Frequency: 4 issues per year
Publication Dates: February, May, August, and November
Language: English
Scope: TESOL, Languages & Linguistics, Education, Pedagogies & Language Teaching, Translation & Interpretation, Cultural studies, Quality Assurance in Education
Article Processing Charges: $70
Types of Journal: Academic/Scholarly Journals
Indexed: Google Scholar, ORCID, Crossref, DOI, Library of Congress, PKP PN, LOCKSS, CLOCKSS, Semantic Scholar, ROAD, EuroPub, J-Gate, Simon Fraser University
Policy: Double blind peer review
Plagiarism check: Turnitin
Contact: editor@i-jte.org; publishing@i-cte.org