TESOL & EDUCATION

INTERNATIONAL JOURNAL OF

TESOL & EDUCATION

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

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

Principal Contact
Associate Professor Dr. Pham Vu Phi Ho
Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam
Email: editor@i-jte.org or ijte.editorial@gmail.com

Reviewers

Table of Content

Research Articles   *Pages*

Literature Review

A Note from the Editor-in-Chief

Dear TESOLers & Educators,

It is with great pleasure that we announce the successful publication of Volume 5, Issue 1, 2025 of the International Journal of TESOL & Education. This issue features a collection of insightful research papers that contribute significantly to the field of TESOL and language education, each offering valuable findings and implications for educators, researchers, and policymakers.

In this issue, Nguyen (2025) investigates the washback effects of IELTS as a high-stakes test on Vietnamese university students' learning. Using a hierarchical model and PLS-SEM analysis of 228 questionnaire responses, the study finds that test utility significantly influences motivation and restricted learning. It highlights the predictive role of test factors in shaping students' psychological and behavioral learning aspects.

Ha and Ho (2025) explore EFL postgraduate students' perceptions of Grammarly and peer feedback in academic writing. Using structured interviews with ten students at Van Lang University, findings reveal that while Grammarly offers quick corrections, peer feedback is more constructive. Most students preferred peer feedback for its depth and relevance, aiding writing improvement.

Pham and Cao (2025) conduct a systematic review of 12 studies to explore the role of ChatGPT in English teaching and learning in Vietnam. The study highlights ChatGPT's benefits in enhancing personalized learning, engagement, and efficiency. However, concerns include over-reliance, academic dishonesty, and the need for training. The review provides recommendations for balanced AI integration in education.

Ngo and Vo (2025) review the application of Augmented Reality (AR) in English Language Teaching (ELT) to support diverse learning styles. Using the Universal Design for Learning (UDL) framework, they highlight AR's benefits for visual, auditory, and kinesthetic learners. The study emphasizes AR's potential for engagement, inclusivity, and accessibility but notes challenges in implementation, accessibility, and teacher readiness.

We extend our heartfelt appreciation to the authors of this issue, whose rigorous research and commitment to advancing TESOL and education have enriched this publication. Their scholarly contributions have provided invaluable insights into current educational trends and innovations, furthering the mission of our journal.

A special note of thanks is extended to our dedicated reviewers, who have generously devoted their time and expertise to ensuring the quality and rigor of the articles published in this issue. Their meticulous feedback and constructive critiques have been instrumental in maintaining the high standards of the International Journal of TESOL & Education. We deeply appreciate their commitment to scholarly excellence.

We also wish to express our sincere gratitude to the editorial staff of the International Journal of TESOL & Education. Their tireless efforts in managing submissions, coordinating peer reviews, and overseeing the publication process have been indispensable in bringing this issue to fruition. Their dedication and professionalism continue to drive the success of our journal.

As we celebrate this accomplishment, we warmly invite researchers, educators, and scholars to submit their original research articles, reviews, and book reviews for consideration in upcoming issues of the journal. We welcome high-quality submissions in all areas related to TESOL, applied linguistics, education technology, and language learning methodologies.

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

IELTS Washback as a High-Stakes Test on Student Learning: A Hierarchical Modelling Study at a Vietnamese University

Nguyen Thanh Minh[1*]

[1] Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam
*Corresponding author's email: minh.nt@vlu.edu.vn
* https://orcid.org/0000-0001-6441-6732
https://doi.org/10.54855/ijte.25511

ABSTRACT

Keywords: IELTS; high-stakes test; test factors; student learning; hierarchical modelling

IELTS has been considered a high-stakes test in the Vietnamese context when its results apply to various purposes. Also, many universities in Vietnam have adopted the IELTS test as the requirement for students to graduate, which may negatively affect their learning. This study aimed to investigate these effects by identifying the mechanism between test factors and the washback of the IELTS test on students' psychological and behavioral aspects of learning. With the involvement of 228 participants, quantitative questionnaires were distributed to collect the data. Then, a hierarchical model was established and analyzed by using the PLS-SEM approach. The findings showed students considered the test utility the most important factor, compared to test stakes and difficulty, because, in the research context, the use of the test received more attention from students. Additionally, within the aspects of learning, learning motivation, and restricted learning were significantly influenced by the washback of the test due to the fact that students were motivated to learn for the test. Finally, the study suggested that test factors played significant roles in predicting the washback of the IELTS test on student learning.

Introduction

Back in the 1980s, when the studies on washback emerged, the focus was mainly on teachers and teaching (Watanabe, 2004). Although students and their learning are directly related to testing and assessment, they are likely to receive less attention from researchers than others (Cheng et al., 2011; Xie & Andrews, 2013). However, washback on students and their learning has recently been paid more attention as a response to address this gap (Sadler, 2016; Xu & Liu, 2018), despite the fact that it remains significant (Cheng et al., 2015). Together, these two directions of research contribute to the descriptions of how to identify washback and reasons for the appearance of the washback concept; nevertheless, these findings could not successfully indicate the mechanism of how washback has on both learning and teaching (Cheng et al., 2011; Xie & Andrews, 2013). One main reason for this issue is the employment of the qualitative approach, which could sufficiently identify factors influencing the washback mechanism but

not fully explain how strong or weak these factors affect the washback mechanism (Xie, 2015). Also, these studies tend to focus on the impacts of test preparation, which means students are in the test preparation courses. This could be misleading because it is difficult to identify the influence of teaching on learning despite the fact that teaching has been considered the most crucial factor in the result of test preparation courses (Zhan & Wan, 2016). Besides, the effect of test preparation is only an example of washback, which could not reveal the complete picture of washback (Dong, 2020). Hence, the quantitative approach is needed to investigate the washback mechanism more by identifying the statistical relationship among different variables as well as between these variables and washback.

What is more, in terms of washback on specific kinds of tests, the literature suggests that high-skate tests, such as IELTS, are in favor of many researchers, which contributes to the knowledge of how washback works in various settings (Tsagari & Cheng, 2017). IELTS is considered one of the most popular English proficiency tests worldwide (IELTS, 2021) because it offers test users and test takers in non-English-speaking countries simple, easily comprehensible, and clear time-bound evidence of an individual's English proficiency (Pearson, 2019). In the Asian context, including Vietnam, IELTS is metaphoric as a "fever" because it is not only a gateway to study or immigrate overseas but also a gateway to graduation and employment (Nguyen & Nguyen, 2022). Due to the appraisal of society for IELTS, many universities have adopted IELTS as a language standard or requirement for students to graduate (Nguyen, 2023). This raises concerns about how IELTS, as a standard for the English level, affects student learning at the university level (Allen, 2016). However, in terms of the test washback on students' learning, these studies mainly focus on student learning behaviors, such as test preparation, test-taking strategies, or the implementation of the test score etc. Consequently, the student learning psychology, such as motivation and anxiety, is not sufficiently addressed (Nguyen, 2023). Along with the washback on learning psychology, the factors that cause this washback primarily rely on contextual, teacher-related, and learner-related but not the test itself (Nguyen, 2023; Watanabe, 2004).

In Vietnam, implementing the National Foreign Language Project, which aims to reform English teaching, learning, and assessment, considerably impacts society (Bui & Nguyen, 2016). For university students, the English proficiency required after graduating is at least B2 level within the Vietnam Framework of Foreign Language Competency (VFFLC), which equals IELTS 6.0 and above. This creates pressure to learn English for all students and even causes a loss in their learning motivation. Also, for the curriculum developers, the issue would be integrating the majors' specific knowledge and the English program so that students could pass both requirements (Albright, 2018; Tran, 2021). Despite the efforts of equipping university students with higher English proficiency, Vietnamese students still have not met the demands of the work market and society's development (Bui & Nguyen, 2016). Additionally, the complex socioeconomic background and the differences in English input of students make the situation even worse (Ehtsham et al., 2023). Considering all of this, the investigation of the effects of IELTS tests as the graduation requirements on student learning is essential.

Literature review

IELTS as a high-stake test

According to its official website, IELTS refers to the International English Language Testing System, an English proficiency test owned by the British Council (BC) in partnership with IDP Education and Cambridge Assessment English. Indeed, IELTS assesses the English proficiency of test takers through its test components, including reading, speaking, listening, and writing. Also, IELTS is available in two formats: General for those who want to immigrate to English-speaking countries and Academic for those who wish to apply for further education or professional opportunities as the requirements of the host countries (IELTS Homepage). Due to its wide application and its effects on test takers, in most context, IELTS is often viewed as high-stake (Clark et al., 2021).

Regarding the high-stake test, there are several definitions in the literature. Johnson (2008) also indicates that a test is considered high-stake when its test score is used as the gatekeeper for passing or failing students, deciding student graduation, examining teachers' accountability, and schools' image and funding. Share the same perspective, Noori and Mirhosseini (2021) indicate that the outcome of high-stake tests could have crucial effects on test takers. Considering all these definitions, in this study, IELTS as a high-stake test is defined as an English proficiency test which has pivotal impacts on learners, teachers and other stakeholders, and it is considered as a gatekeeper for school admission, graduation, job prospects, or immigrating (Johnson, 2008; Noori & Mirhosseini, 2021).

Effects of testing as impact, washback, and consequence

There are different terms used to describe the impacts of testing, especially high-stake testing, as an educational phenomenon when its results significantly affect the stakeholders related to such testing, including impact, washback, and consequence (Tsagari & Cheng, 2017). Specifically, washback is often used to indicate the effects of testing on learning and teaching in the classroom context (Hughes, 2002). Meanwhile, impact refers to a broader view of the effects of testing as any effects that have on test-takers, policies, or practices within and/or beyond classroom settings such as schools, educational systems, or even society (Wall, 2012). As a result, many researchers and language testers consider washback as one dimension of the impacts of testing (Hamp-Lyons, 1997). What is more, the effects of testing on teaching and learning are usually related to test validity, in which washback needs to be taken into account in measuring test validity as a testing consequence (Messick, 1996). In this study, the effects of testing are examined in relation to student learning; hence, the term washback is employed as the shortened washback of testing.

Natures of washback

Washback is usually identified within five dimensions: specificity, intensity, length, intentionality, and value (Cheng & Watanabe, 2004). *Specificity* refers to the level of specificity of the washback; in other words, washback could be general or specific. Indeed, general washback indicates the effects created by any kind of test, while specific washback presents the impact of a certain type of test or a particular aspect of a test (Alderson & Wall, 1993; Cheng & Watanabe, 2004). *Intensity* describes the power of washback, i.e. the degree of effects in areas of teaching and learning produced by a test (Cheng, 1997). Generally, a high-stake test would affect teaching and learning more than a classroom-based test (Xu & Liu, 2018). *Length* of a washback indicates how long a test's effects lasts (Watanabe, 2004). *Intentionality* of washback means that washback could be intended or unintended. To be more specific, intended washback refers to effects which test designers expect. In contrast, unintended washback

illustrates unexpected effects, such as student anxiety or long-term training on taking tests (Xu & Liu, 2018). Finally, *value* or direction of washback asserts that washback could be positive or negative. For instance, a well-designed and appropriate test could be beneficial to teaching and learning by offering helpful information and creating student motivation. However, a test that is not well-developed or suitable would create negative effects (Xu & Liu, 2018).

Washback models on student learning

Although many conceptual models of washback have been developed, washback studies have indicated that significant variability exists in how teachers, learners, and even other stakeholders have adjusted their behaviors and attitudes towards different testing demands (Ha, 2019). This leads to researchers' struggling to find a common framework to capture this dynamic nature and all the variations of washback (Green, 2013; Liu & Yu, 2021). However, two washback models are repeatedly used by many studies in the field, including the washback trichotomy model by Hughes (1993) and the washback model of student learning by Shih (2007).

The model of Hughes (1993) presents the trichotomy into three main factors, including participants, process and products to construct the basic model of washback. In this model, participants refer to anyone whose work is affected by the perceptions and attitudes towards a test. The process indicates the actions contributing to the learning process, and the product presents what has been learned or acquired (e.g. skills, facts, etc.) and the quality of learning (Hughes, 1993). Also, the mechanism among these is clarified. Testing, first and foremost, affects the participants' perceptions (e.g. teachers and learners) towards their teaching and learning tasks. These perceptions, in turn, impact how these participants conduct their work (process), such as test preparation on similar test items, which will influence the learning outcomes (the products of this work) (Hughes, 1993; Xu & Liu, 2018). In brief, in the model of Hughes (1993), learning behaviors are emphasized.

Regarding the washback model of student learning, Shih (2007) states that the basic model mentioned above could not fully cover washback in social and educational contexts due to the variability of individual learners. Hence, she proposes a model that considers the effects of washback on learning in different psychological aspects, including students' thoughts, experiences, and feelings. She also describes how washback affects students' psychology via extrinsic, intrinsic, and test factors (Shih, 2007).

By examining these two models of washback, this study employs a combination of approaches proposed by Hughes (1993) and Shih (2007), i.e. examining washback on learning in both behavioral and psychological aspects of student learning.

Test factors of IELTS as a high-stake test

Several factors generate washback of a test, including factors related to context, teachers, learners and the test itself (Watanabe, 2004). These proposed factors have been repeatedly examined in different studies. Therefore, in the current study, the test factors are taken into account because there is a limited number of empirical studies on how the test factors influence the washback of the test on student learning (Nguyen, 2023). There are several perspectives on the components of test factors. Watanabe (2004) proposes five components, including test methods, test content, skills tested, test purpose, test stakes and test status. However, these components partly influence the washback; hence, Shih (2007) and Xie (2015) consider test difficulty to directly affect student learning since students' expectations to pass the test according to its difficulty would enhance their learning. Besides, the implication of the test in reality also impacts students' learning. Specifically, it is often regarded as the extrinsic

motivation for student learning (Xie & Andrews, 2013). Bachman (2005) states that the use of tests could be conceptualized as test utility, i.e. the interpretation of the test score and the use of it in learners' situations. Finally, the test had different impacts on learners when it is known as a high-stake or low-stake test. For instance, in the context of not being recognized in the academic transcription, IELTS is viewed as a low-stake test that does not govern students' learning (Stoneman, 2006, as cited in Tsang and Isaacs, 2022). Meanwhile, if the IELTS test plays a role as a graduation gateway, the washback of it is more intensive (Allen, 2016). Based on these reviews, this study viewed test factors as a combination of three aspects: test difficulty, test utility, and test stakes.

Current research on the washback of student learning

The literature has revealed that the effects of washback on student learning are "mixed" between positive and negative (Ha, 2019). The sections below outline the current findings on washback's positive and negative effects.

Positive washback

Some repeatedly cited studies by Xiaoju (1990), Cheng (1998), Hirai and Koizumi (2009), Pan and Newfields (2011), and Allen (2016) show positive washback on student *learning motivation* toward English. Motivation is often considered an essential factor in language learning because it governs and helps maintain students' learning activities as well as influences student engagement and attainment (Schunk, 2012). According to Dörnyei and Ushioda (2021), motivation could be clarified into intrinsic and extrinsic motivation. A test could create intrinsic motivation that students tend to request more extra English materials, participate in activities and read more English materials or journals as well as watch English TV programs (Hirai & Koizumi, 2009; Xiaoju, 1990). Students also spend more time learning English to be better prepared for the test as the results of extrinsic motivation, i.e. to pass the test (Allen, 2016; Cheng, 1998; Pan & Newfields, 2011) and the test result could possibly affect their future prospects (Chu & Yeh, 2017; Nhan, 2013). However, studies by Cheng (1998), Shih (2007) and Pan and Newfields (2012) reveal that tests have minimal effects on learning motivation.

Another positive impact of washback is on student *holistic learning*, i.e., improving skills and abilities or any encouragement to do so due to the test (Alderson & Wall, 1993; Cheng, 1998). indicates that in China's university context, students employ better-coping strategies, such as test-taking, test management and other meta-cognitive strategies to do the test better. Additionally, attempts to do the tests enhance students' abilities and competence significantly (Hung, 2012).

Negative washback

However, washback, at the same time, creates negative effects on student learning. While washback motivates students to learn, it is also responsible for student *learning anxiety* (Shih, 2007). This is a psychological condition occurring when students lack self-confidence in terms of their competence as well as their test results (Al Hadhrami et al., 2024; Tae-Young & Yoon-Kyoung, 2016). This issue could lead to an adverse impact on their academic performance and to higher levels of stress and depression (Shamsuddin et al., 2013).

Also, washback of testing could lead to *restricted learning*, including rote learning, memorization, past tests, reviewing teachers' notes, etc. (Damankesh & Babaii, 2015; Dong, 2020). Also, students are more likely to employ coping strategies related to the tests rather than cognitive or metacognitive strategies (Xiao, 2014). These lead to some memorization of materials, which could be beneficial to doing the test but not useful for real-life situations or usages (Ren, 2011).

Washback of IELTS on student learning

As mentioned earlier in this section, IELTS is viewed as a kind of high-stake test which greatly impacts several stakeholders, especially students (Johnson, 2008; Madaus & Keillor, 1988; Noori & Mirhosseini, 2021). In the educational context, IELTS plays a role as a tool for measuring students' language ability for enrolment in the academic environment (Pearson, 2019). Due to this nature, in the international context, many researchers have investigated the washback of this test in university settings.

Green (2007) investigated the washback of the IELTS test on students' learning outcomes in terms of their preparation for the test. With a total of 476 participants who took different types of IELTS test preparation courses, the IELTS writing tests and questionnaires were distributed. After using the neural network approach to analyze the data, the author concluded that the course preparation did not affect the scores of the participants. Still, the test difficulty was the determinant of the learners' strategies to prepare for the test.

In another attempt, Stoneman (2006, as cited in Tsang and Isaacs, 2022) compared the impacts of the Graduating Students' Language Proficiency Assessment (GSLPA) and the IELTS on student learning in a Hongkong University. Using a survey and the semi-structured interview as instruments, the study found that these tests significantly affected student motivation to take the preparation course. This could be explained by the awareness of the stakes of these two tests, i.e., an international proficiency exam and a nationwide test. Also, because the IELTS test, in this case, was not included in the transcription, the participants considered its stake to be lower than the other.

Allen (2016) also focused his research on examining the washback of the IELTS test on student test preparation. In his study, 190 participants were involved by completing two IELTS tests, a survey, and an interview for 19 participants. The findings indicated a positive washback in student learning, i.e., improving students' language proficiency with an unequal distribution among the four language skills. This result also varied from those with prior experience in doing the tests and those with high- and low levels of English proficiency. Similar results were also found in the studies of Read and Hayes (2003) and Zhengdong (2009).

In the Vietnamese context, Nguyen (2023) conducted a study on the washback of IELTS on last year English majors' learning at a university. He employed the quantitative research design that followed a combination of frameworks by Hughes (1993) and Shih (2007). Specifically, 282 students participated in his study to respond to a survey on the factors of the IELTS test and how this test affects their learning psychology and behaviors. The collected data were analyzed by using Exploratory Factor Analysis and Confirmatory Factor Analysis with descriptive statistics, Pearson correlation, and multi-linear regression. The findings presented that the IELTS test positively and negatively affected student learning. While IELTS was perceived as important to their study and future lives, participants considered it hard to reach the required score. Also, the study showed that the mechanism of washback via test factors varied in the research context. Besides these contributions to the knowledge of washback, this study validated a survey for measuring washback of IELTS via psychological and behavioral aspects of student learning, which is adapted in the current study.

From the current literature, the available studies that examined the mechanism of IELTS washback on learning in terms of the test factors remain somewhat inadequate. Also, test factors, which have more influence on student learning psychology and behaviors, still have not received enough attention, especially in the Vietnamese context (Nguyen, 2023). Therefore,

this study aimed to examine this mechanism using the hierarchical modelling approach to gain more insights and contribute to the literature on washback studies.

Conceptual framework of the study

From all the reviews above, this study examined the relationship between test factors and the washback of the IELTS test on student learning. Specifically, the test factors are defined into three components: test difficulty, test stakes, and test utility (Allen, 2016; Bachman, 2005; Shih, 2007; Xie, 2015). Also, the IELTS washback on learning are manifested via learning motivation (Cheng, 1998; Hirai & Koizumi, 2009; Pan & Newfields, 2011; Xiaoju, 1990), learning anxiety (Shamsuddin et al., 2013; Shih, 2007), holistic learning (Alderson & Wall, 1993; Cheng, 1998), and restricted learning (Damankesh & Babaii, 2015; Dong, 2020). Figure 1 depicts relationships among these constructs. Additionally, this model has been validated by Nguyen (2023) by using EFA and CFA approaches.

Figure 1

Conceptual framework of the study

Research Questions

Based on the aim stated above and the conceptual framework, three research questions (RQs) have been formulated as follows:

RQ1: How do English majors perceive the factors of the IELTS test in terms of its difficulties, utility, and stakes?

RQ2: How do English majors perceive the impact of the washback of the IELTS test on their learning in terms of their learning motivation, learning anxiety, holistic learning, and restricted learning?

RQ3: Which factors of the IELTS are determinants of the washback of student learning?

Methods

Pedagogical Setting & Participants

The current study was conducted at Vision University (VU, pseudonym), which follows the English requirement of IELTS 6.0 and above for student graduation. The participants in this study were recruited through the convenience sampling method. This kind of non-probability sampling method allows researchers to flexibly contact the participants within their neighborhood or via the Internet (Edgar et al., 2017). In particular, there were 228 participants involved in the study, and the demographic information was presented in the "Findings and Discussion" section.

Design of the Study

This study employed the quantitative approach because the research design provides non-biased and precise measurements, accurately identifying relationship between variables (Mesly, 2015). Also, this design increases the chances of generalizing the results for a larger population (Queirós et al., 2017). The employment of this approach is suitable for the study to investigate the washback mechanism between test factors and student learning.

Research instruments

By conducting a quantitative study, the researcher utilized the questionnaires in collecting data. According to Roopa and Rani (2012), questionnaires allow researchers to collect data from a relatively great number of participants within a short time. Additionally, it offers a more flexible form of distributing via the Internet and in the analysis by using different types of statistical estimation (Creswell, 2014). In this study, the questionnaire was adapted from Nguyen's study (Nguyen, 2023). Table 1 summarizes the constructs and items covered in the instrument.

Table 1

Constructs included in the questionnaire

No.	Constructs		Number of items
1	Demographic information		2
2		Test difficulty	9
3	Test factors	Test stakes	9
4		Test utility	9
5		Learning motivation	10
6	Washback on student learning	Learning anxiety	5
7		Holistic learning	5
8		Restricted learning	5

In short, there were 55 items in the questionnaire, which were divided into three big constructs and 8 sub-constructs: (1) demographic information (gender and year of study), (2) test factors

(test difficulty, test stakes, and test utility), and (3) Washback on student learning (learning motivation, learning anxiety, holistic learning, and restricted learning).

Data collection & analysis

The questionnaire was distributed online to the participants via the invitation of the researchers. After the data collection ended, there were 228 responses to the questionnaire, which were collected and coded for data analysis. In the current study, the data analysis procedures followed the PLS-SEM approach and were performed on the SmartPLS 3.0 software. Hair et al. (2013) and Sarstedt et al. (2017) consider PLS-SEM as an effective way of exploring the relationship between exogenous and endogenous variables via the assessment of measurement and structural models with high accuracy. Under this approach, there were three main stages, including (1) establishing the models, (2) assessing measurement models, and (3) assessing the structural model. The detailed descriptions of this analysis procedures were outlined in the "Findings and Discussion" section.

Validity and reliability

Some techniques were employed to ensure the validity and reliability of the study. Firstly, the conceptual framework and the questionnaire were adapted from Nguyen's study (Nguyen, 2023). Moreover, the questionnaire was piloted to 50 respondents before being distributed to the participants. In the pilot process, the reliability of the questionnaire was assessed, using Cronbach Alpha and Composite Reliability estimations. Hair et al. (2013) assert that the reliability of the questionnaire could be assessed via the measurement of Cronbach Alpha (with the value > 0.05) and Composite Reliability (with a value > 0.708). As Table 3.2 shows, all figures for these two estimations met the requirements. As a result, all the items of the questionnaires are reliable.

Table 2

The estimations of Cronbach Alpha and Composite Reliability for the questionnaire pilot

No.	Items	Cronbach Alpha	Composite Reliability
1	Test factors	0.946	0.953
2	Test difficulty	0.911	0.924
3	Test stakes	0.895	0.912
4	Test utility	0.921	0.930
5	Washback on student learning	0.938	0.941
6	Learning motivation	0.894	0.899
7	Learning anxiety	0.848	0.850
8	Holistic learning	0.920	0.922
9	Restricted learning	0.925	0.930

Results/Findings and discussion

Descriptive statistics

In the current study, 228 participants responded to the online survey. The following figures illustrate their demographic information, including gender and their school years.

Figure 2

Gender distribution of the participants
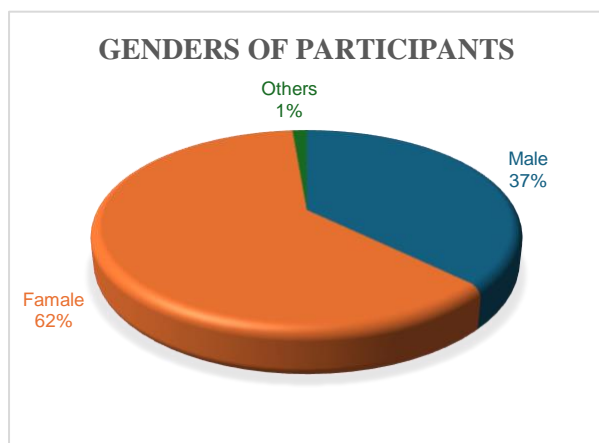
Figure 3

Distribution of school years of participants
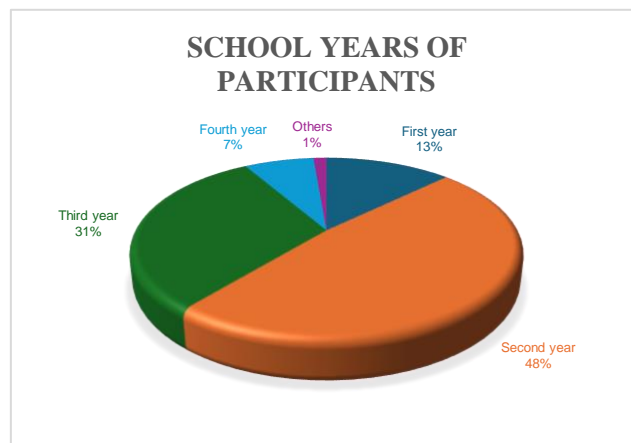
Among the participants, the majority of them were female students, with a percentage of 62%, followed by the figures of males (37%) and others (1%). Regarding their years of study, most of them were in their second year (48%) and third year (31%).

Establishing and assessing the hierarchical model

According to the hypotheses and the conceptual framework proposed after reviewing the current literature, the hierarchical data analysis model was employed to present the relationship between test factors and washback on student learning. Generally, hierarchical models or multidimensional models are considered when researchers aim to investigate constructs with more than one dimension (Crocetta et al., 2021; Heck & Thomas, 2020). In the current study, the latent variable (LV) "test factors" is defined via three manifest variables (MVs), including test difficulty, test stakes, and test utility. Similarly, the latent variable (LV), "washback on student learning", is defined via four manifest variables (MVs), including learning motivation, learning anxiety, holistic learning, and restricted learning. Therefore, two dimensions are involved in the study, forming the hierarchical model's first and second order. This kind of model is beneficial in decreasing the complexity of the model and is more sufficient to utilize the available resources (Crocetta et al., 2021; Heck & Thomas, 2020). Also, the hierarchical model provides high measurement validity (Law et al., 1998).

The PLS-SEM approach is often used to analyze hierarchical models because it offers a tool for examining the relationship and influence of different aspects of a phenomenon (Crocetta et al., 2021). Additionally, the PLS-SEM approach allows the conceptualization of hierarchical models by the repetition of LVs within the model (Guinot et al., 2001). To establish the model using PLS-SEM, this study followed the guidelines from Wetzels et al. (2009) as presented in Figure 4.

Figure 4

Steps to establishing the hierarchical models by using the PLS-SEM approach (adapted from Wetzels et al. (2009))
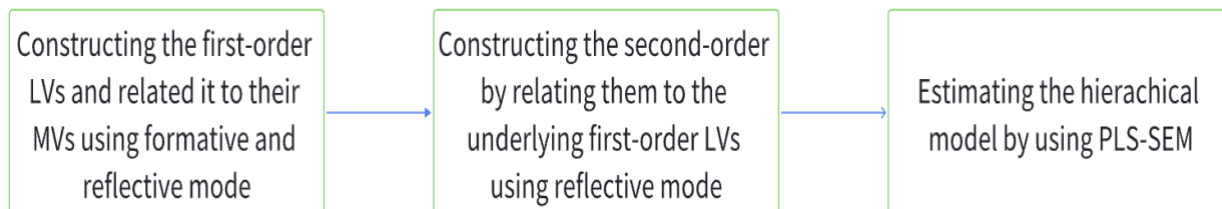
As Figure 4 suggests, the first step is constructing the first-order LVs and relating them to their MVs using reflective and formative modes. In this study, the first-order models included two sets of MVs, which consisted of seven MVs: test difficulty, test stakes, test utility, learning motivation, learning anxiety, holistic learning, and restricted learning. Then, in the next step, the second-order models of LVs were constructed by relating two LVs – test factors and washback on student learning with their underlying MVs. Specifically, the LV "test factors" was determined by three MVs - test difficulty, test stakes, and test utility – under the formative model. Concerning the LV "washback on student learning", it was manifested by four MVs - learning motivation, learning anxiety, holistic learning, and restricted learning – under the reflective model. After identifying the first and second-order models, the final measurement hierarchical model was proposed by using the PLS-SEM approach, as illustrated in Figure 5. To easily import to the software for analysis, the indicators of all the variables were coded in Table 3.

Table 3

Codes used in the data analysis procedures

No.	Constructs		Codes of indicators/ items
1	Gender		Gen
2		Test difficulty	TestDif1→ TestDif9
3	Test factors	Test stakes	TestStake1 → TestStake9
4		Test utility	TestUtility1 → TestUtility9
5		Learning motivation	LearnMoti1 → LearnMoti10
6	Washback on student learning	Learning anxiety	LearnAnxi1 → LearnAnxi5
7		Holistic learning	HoLearn1 → HoLearn5
8		Restricted learning	ResLearn1 → ResLearn5

Figure 5

The hierarchical model of the study
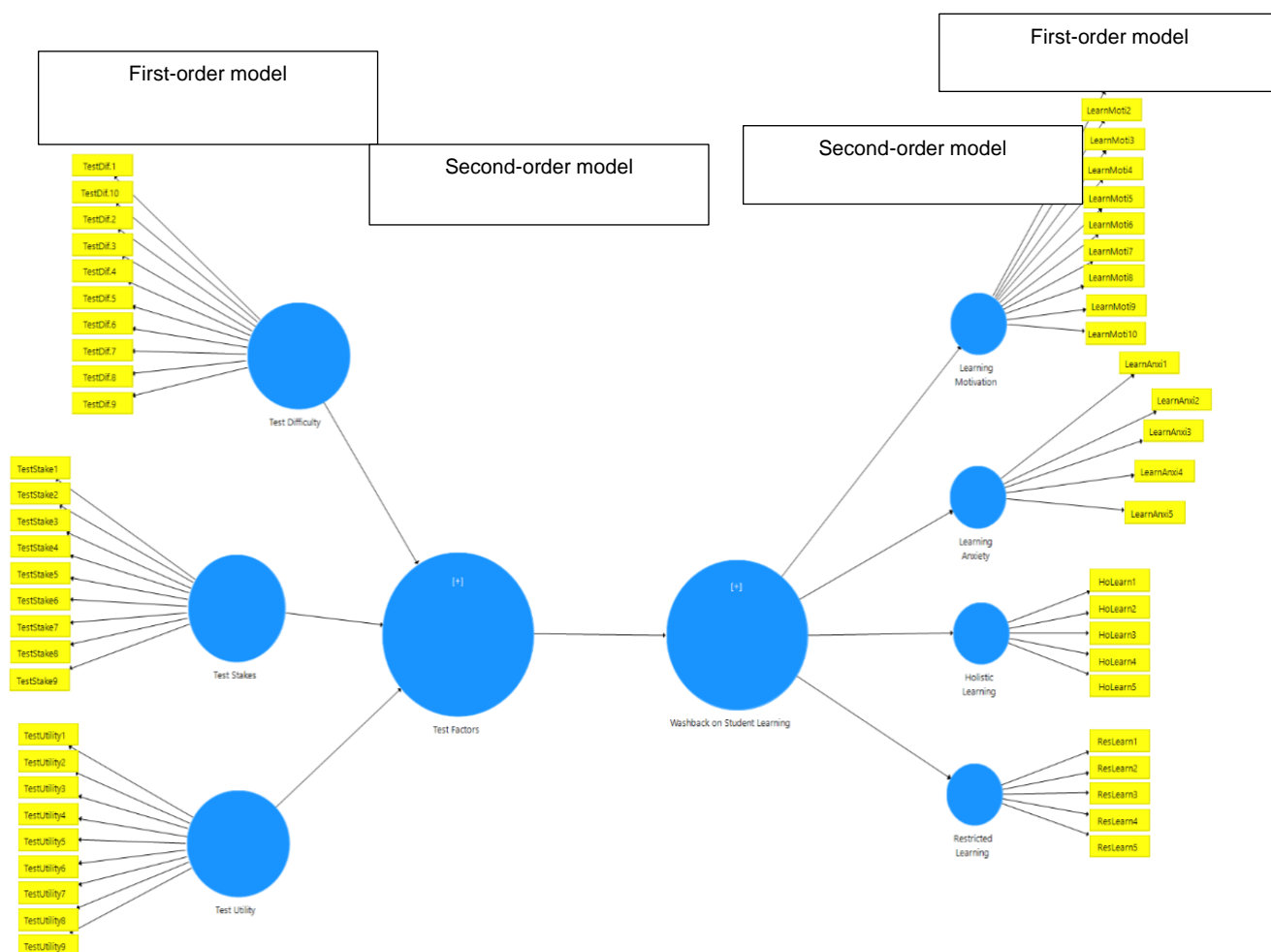
Also, in the final step, the final model was estimated using the PLS-SEM approach in the SmartPLS 3.0. This process was performed within two stages: (1) assessing the measurement model via all the psychometric values of first-order models and (2) assessing the structural model, i.e. the magnitude of the relationships or effects between the variables being considered within the model, including second-order model in Figure 5 (Marcoulides & Saunders, 2006).

Assessing the measurement models

To assess all psychometric values of all nine variables in the model, the outer factor loading, the composite reliability (CR), convergence validity (AVE), and discrimination validity were taken into account (Hair et al., 2013). Firstly, the outer loading factors were estimated to eliminate the indicators that had the value smaller than 0.7 as these did not measure the construct they were supposed to measure within the study's sample. Within this process, the indicators "TestDif3", "TesDif9", and "TestStake9" were removed because in the current estimation, the values of these indicators were below 0.7. As a result, the model with the loading factors is shown in Figure 6.

Figure 6

Hierarchical model with path coefficient
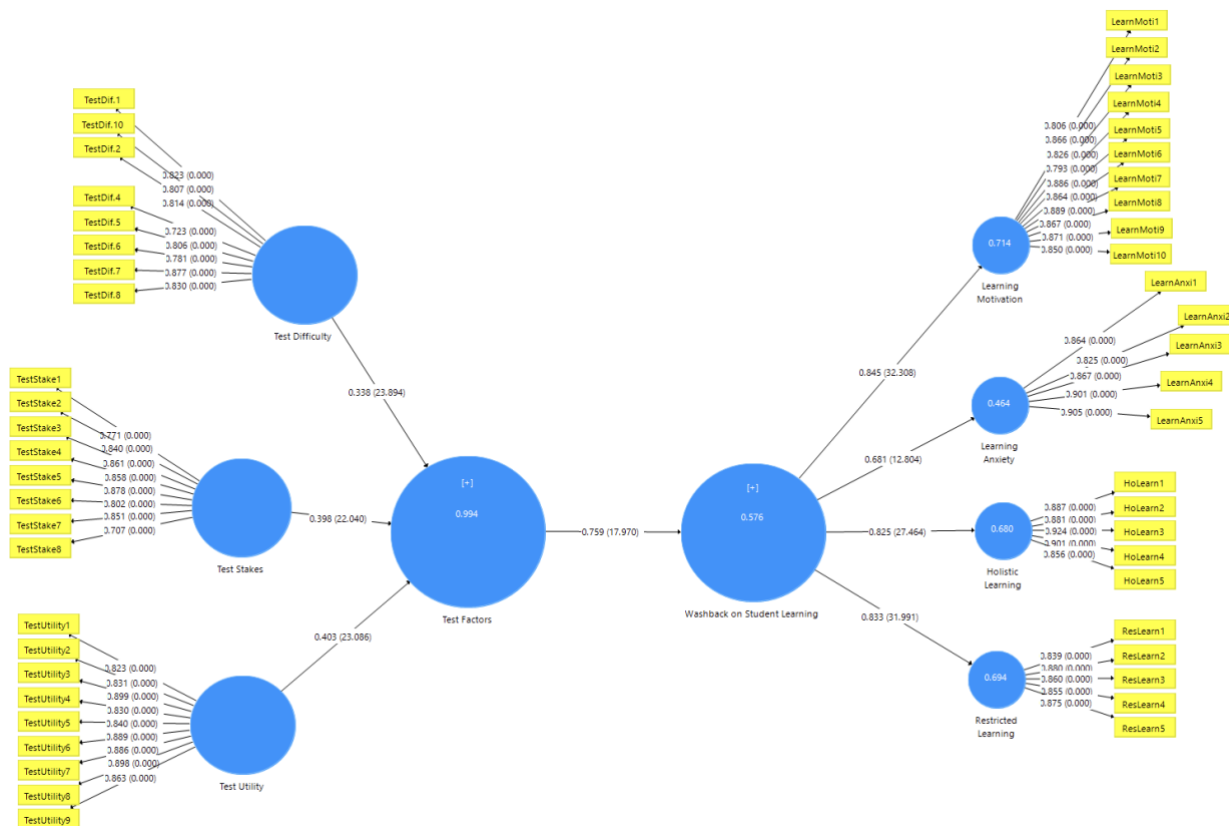
After the removal of all unsatisfied indicators, the remaining values – the composite reliability (CR), convergence validity (AVE), and discrimination validity – were assessed.

Table 4

The psychometric values of the first-order models

Construct	CR	AVE
Test difficulty	0.938	0.654
Test stakes	0.943	0.677
Test utility	0.963	0.744
Learning motivation	0.964	0.726
Learning anxiety	0.941	0.762
Holistic learning	0.950	0.792
Restricted learning	0.935	0.743

In the model, the CR values of the variables in the model must be above 0.78, and AVE must be above 0.5 to ensure that the estimation of the constructs and their relationships are objectives

and not impacted by measurement errors (Hair et al., 2019). According to Table 4, all the values of CR and AVE of the variable in the first-order model satisfy this requirement.

Table 5

The HTMT matrix of the first-order models

	Holistic Learning	**Learning Anxiety**	**Learning Motivation**	**Restricted Learning**	**Test Difficulty**	**Test Stakes**
Learning Anxiety	0.353					
Learning Motivation	0.755	0.397				
Restricted Learning	0.58	0.613	0.585			
Test Difficulty	0.469	0.55	0.462	0.469		
Test Stakes	0.644	0.473	0.719	0.555	0.607	
Test Utility	0.682	0.418	0.699	0.558	0.581	0.846

In terms of discrimination validity, the HTMT matrix was examined. As Henseler et al. (2015) suggest, all values of the HTMT matrix should be under 0.78 to prove that the examined constructs do not overlap. From Table 5, all the values met the standards.

Assessing the structural model

After ensuring that the first-order models were all valid, Hair et al. (2013) and Sharma and Aggarwal (2019) propose the following statistical estimation should be performed to assess the relationship of the whole model: Collinearity, R^2 explanation of endogenous latent variables, f^2 effects size of path coefficients, and Predictive relevance Q^2. Also, at this stage, the nonparametric Bootstrapping was used with 5,000 replications to examine the hypothesis with a significance of 0.05.

Firstly, for collinearity, the VIF was estimated with a standard smaller than the threshold of 3.3 (Roberts & Thatcher, 2009). Via the examination of VIF, the maximum value of the current model was 2.967, which met the standard above. Therefore, in the model, there was no risk of collinearity among variables.

The estimation of R^2 *and* R^2 adjusted are important in determining to what extent the MVs could explain the LVs or in-sample predictive power (Nguyen & Vu, 2020). In the model of the study, the variable Test factors affected the variable Washback on student learning, while the four variables – Holistic learning, Learning anxiety, Learning motivation, and Restricted learning – reflected the variable Washback on student learning. The powers of explanation for these variables are presented in Table 6. According to Henseler et al. (2015), there are three degrees of effect, including low (<0.25), moderate (> 0.25 and < 0.50), and high (>75%).

Table 6

The estimation of R^2 and R^2 adjusted

	R Square	R Square Adjusted	Comment
Holistic Learning	0.68	0.679	Moderate effect
Learning Anxiety	0.464	0.461	Moderate effect
Learning Motivation	0.714	0.712	Moderate effect
Restricted Learning	0.694	0.692	Moderate effect
Test Factors	0.994	0.994	High effect
Washback on Student Learning	0.576	0.574	Moderate effect

As Table 6 indicates, most of the variables had moderate effect in in-sample predictive power, except test factors, which could explain Washback on student learning at a high effect degree.

Turning to f²effects size of path coefficients, the results are shown in Table 7.

Table 7

The estimation of the f²effects size of path coefficients

	P Values
Test Difficulty → Test Factors	0.000
Test Factors → Washback on Student Learning	0.000
Test Stakes → Test Factors	0.000
Test Utility → Test Factors	0.000
Washback on Student Learning → Holistic Learning	0.000
Washback on Student Learning → Learning Anxiety	0.001
Washback on Student Learning → Learning Motivation	0.000
Washback on Student Learning → Restricted Learning	0.000

In Table 7, all the p-values of the f²effects size of path coefficients were smaller than 0.05. As a result, all the MVs explained the LVs.

Lastly, the Predictive relevance Q^2 was estimated to identify the model's predictive power via the Blindfolding process.

Table 8

The estimation of the Predictive relevance Q^2

	Q^2 (=1-SSE/SSO)
Holistic Learning	0.532
Learning Anxiety	0.345
Learning Motivation	0.509
Restricted Learning	0.509
Washback on Student Learning	0.357

There are two standards for evaluating the Predictive relevance Q^2, including the value of Q^2 and the range of Q^2 value. First of all, if all the Q^2 values of all variables are above 0, then the structural or Hierarchical models in this study reach the global quality (Tenenhaus et al., 2005). In addition, the predictive power of the model is ranked into three levels, including low (the value is between 0 and 0.25), moderate (the value is from 0.25 to 0.5), and high (the value is above 0.5). By examining Table 8, all the figures were nearly over 0.5, which was considered as high predictive power, except the figures for Learning Anxiety and Washback on Student Learning (with moderate predictive power).

Discussion

From the estimation of the current model within the study sample, it is concluded that the proposed model was valid in terms of the measurement of each MV and LV as well as their relationships. The next step was to address the research questions with these findings.

RQ1: How do English majors perceive the factors of the IELTS test in terms of its difficulty, utility, and stakes?

The following table summarizes the result of the model related to the relationship between test difficulty, test stake, test utility and test factors.

Table 9

Total effects of factors of IELTS test

Relationship	Path coefficient	P-value
Test difficulty → Test factors	0.338	0.000
Test stakes → Test factors	0.398	0.000
Test utility → Test factors	0.403	0.000

According to Table 9, the participants in the study considered test difficulty, test stakes and test utility to be highly correlated to test factors (p-value < 0.05). Test utility was also regarded as the most influential factor among these three factors, followed by test skates and test difficulty (the value of path coefficient of 0.403, 0.398, and 0.338, respectively). In the research context,

in Vietnamese universities, the IELTS test is highly regarded; however, it is not included in the academic transcriptions. This is in contrast to what Shih (2007) and Xie (2015) found in their research, which found that test difficulty is the most influential factor. Hence, the test stakes in the study did not receive the highest weighting, compared to the utility, which aligns with the finding of the study conducted by Stoneman (2006, as cited in Tsang and Isaacs, 2022) that the IELTS test was not perceived as important when its stakes were low.

RQ2: How do English majors perceive the impact of the washback of the IELTS test on their learning in terms of their learning motivation, learning anxiety, holistic learning, and restricted learning?

Table 10 presents the relationship between the washback of the IELTS test and different facets of student learning.

Table 10

Total effects of washback on learning of the IELTS test

Relationship	Path coefficient	P-value
Washback on student learning → Learning motivation	0.845	0.000
Washback on student learning → Learning anxiety	0.681	0.000
Washback on student learning → Holistic learning	0.825	0.000
Washback on student learning → Restricted learning	0.833	0.000

As Table 10 suggests, the washback of the IELTS test had a positive relationship with students' learning aspects, including learning motivation, learning anxiety, holistic learning, and restricted learning (p-value <0.05 and values of coefficient> 0). These results clearly illustrated the fact that washback of the test statistically significantly influences learning behaviors and psychology of student learning. This is confirmed in the study of Nguyen (2023) when both psychological and behavioral aspects of student learning were considered. Regarding each aspect in isolation, the effects of the test on students' learning motivation were found by the research of Xiaoju (1990), Cheng (1998), Hirai and Koizumi (2009), and Pan and Newfields (2011), and Allen (2016), while these impacts on students' holistic learning were also identified by Alderson and Wall (1993), Cheng (1998) and Xiao (2014). Specifically, the washback of the test increased learners' motivation to learn English, and in this study, the washback of the IELTS test highly affected learning motivation (with a coefficient of 0.845). This finding was on the contrary with the minimal influence of the test washback and motivation the research by Cheng (1998), Shih (2007), and Pan and Newfields (2012).

Holistic learning was viewed as the positive aspect of learning that the test had washback on (Alderson & Wall, 1993; Cheng, 1998). In this study, the participants perceived this aspect in a lower manner than that of restricted learning, which shared the same results as the study of Xiao (2014), Damankesh and Babaii (2015), and Dong (2020) when learners focused more on the strategies to cope with the test. However, the findings still indicated that learners saw that their English skills improved during the test preparation. Another negative effect of washback on learning is learning anxiety (Shih, 2007). The current study revealed that the test did not correlate with anxiety as highly as the other learning aspects. This finding is in contrast with the studies of Shamsuddin et al. (2013) when the anxiety of students was found to be significant.

RQ3: Which factors of the IELTS are determinants of the washback of student learning?

Table 11 shows the effects of test factors on the test washback of students' learning.

Table 11

Total effects of test factors on the test washback of students' learning.

Independent variable	Dependent variable	Path coefficient	P-value
Test difficulty	Washback on student learning	0.255	0.000
Test stake	Washback on student learning	0.303	0.000
Test utility	Washback on student learning	0.305	0.000

According to Table 11, the components of test factors, including test difficulty, test stakes, and test utility, had statistically significant effects on students learning, known as test washback (path coefficient> 0, and p-value < 0.05). Within the study samples, the test utility had the most influence on the test washback on learning, followed by the test stakes and test difficulty. For the participants of this study, the test usages play the most significant role in their learning English. This reflects the assumption of Pearson (2019) and the social context of Vietnam (ThanhNienNews, 2022) that the IELTS test provides a tool for measuring English proficiency for various purposes, from academic to employment. Also, test stakes impacted student learning in the second position with just a small loading factor of 0.02. This aligned with the conclusion of Stoneman (2006, as cited in Tsang and Isaacs, 2022) and Allen (2016) that the stakes of a test are considered as high when its usages are significant and directly affect the test taker. However, the test difficulty did not affect the students' learning as much as the other. This could be explained by the common knowledge and awareness of the test taker when the IELTS test becomes dominant within international English proficiency tests (IELTS, 2021). Compared to the study of Nguyen (2023), this study found that all test factors had considerable influences on learning aspects, not just some aspects such as learning motivation or restricted learning.

In addition to these findings, the current study contributed to the validation of the washback model based on test factors. The estimation of Predictive relevance Q^2 in the above section indicates that the test factors strongly predicted the impact of the IELTS test on students' holistic learning, learning motivation and restricted learning (Q^2 value > 0.5). Hence, test factors should be considered as an important determinant of the test washback, as suggested by Shih (2007), Xie (2015) and Nguyen (2023).

Conclusion

The current study aimed at investigating the mechanism of how test factors of the IELTS test (including test difficulty, test stakes, and test utility) affect the washback on student learning in terms of learning motivation, learning anxiety, holistic learning, and restricted learning. With the involvement of 228 students at Vision University (VU) and the employment of a survey and PLS-SEM approach in analyzing the hierarchical model, the study indicated that test factors played a significant role in the mechanism of washback on student learning. Specifically, among three test factors, the learners in the research context considered test utility the most important one, followed by test stakes and difficulty. Regarding student learning, learning

motivation and restricted learning were recorded as the most affected aspects by the test washback. Regarding the relationship between test factors and test washback on student learning, test utility and test stakes were the determinants compared to test difficulty. The hierarchical model proposed in the study yielded the predictive power of the models with the examination of test factors, and both psychological and behavioral aspects of learning was relatively high. From the results of the current study, it is apparent that in the context of Vietnam and some other familiar contexts in Asia, the IELTS test is used for various purposes, which indicates the importance of its utility. However, this test is not included in students' transcription; therefore, the stakes of it is low. Additionally, due to its popularity among students, the test format is somewhat familiar to them. As a result, among the test factors, test difficulty is not viewed as the most influential. Finally, the study confirms the possible impacts of high-stake tests on both psychological and behavioral aspects of student learning.

Although the study contributes to the understanding of the washback mechanism of test factors on both learners' psychological and behavioral aspects of learning, there still remained some limitations. First of all, the sample size and sampling methods employed in the study could reduce the generalization of the findings. However, in the study, utilizing the bootstrapping technique to analyze hierarchical models could minimize that risk. Therefore, it is recommended that other researchers could use probability sampling methods with a larger sample size for their studies. Additionally, due to the time constraint, the research design relied on quantitative research design. Nevertheless, as stated in the introduction section, most studies on washback of testing on learning employ qualitative research design. Hence, this study attempted to bridge this gap.

References

Al Hadhrami, I., Al Sawafi, A., Abraham, J., Al Aamri, K., & ALzeidi, A. (2024). Test anxiety: perceptions of Omani nursing students–a descriptive qualitative study. *Nurse Education in Practice*, *76*(1), 10328-10335. https://doi.org/10.1016/j.nepr.2024.103928

Albright, J. (2018). *English tertiary education in Vietnam*. Routledge.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, *14*(2), 115-129. https://doi.org/10.1093/applin/14.2.115

Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, *6*(1), 7. https://doi.org/10.1186/s40468-016-0030-z

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, *2*(1), 1-34. https://doi.org/10.1207/s15434311laq0201_1

Bui, T. T. N., & Nguyen, H. T. M. (2016). Standardizing English for Educational and Socio-economic Betterment- A Critical Analysis of English Language Policy Reforms in Vietnam. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 363-388). Springer International Publishing. https://doi.org/10.1007/978-3-319-22464-0_17

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and education*, *11*(1), 38-54. https://doi.org/10.1080/09500789708666717

Cheng, L. (1998). Impact of a Public English Examination Change on Students' Perceptions and Attitudes toward Their English Learning. *Studies in Educational Evaluation*, *24*(3),

279-301. https://doi.org/10.1016/S0191-491X(98)00018-2

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, *28*(2), 221-249. https://doi.org/10.1177/0265532210384253

Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, *48*(4), 436-470. https://doi.org/10.1017/S0261444815000233

Cheng, L., & Watanabe, Y. (2004). *Washback in language testing: Research contexts and methods*. Routledge.

Chu, H.-y., & Yeh, H.-n. (2017). English Benchmark Policy for Graduation in Taiwan's Higher Education: Investigation and Reflection. *Journal of Language Teaching and Research*, *8*(6), 1063-1072. https://doi.org/10.17507/jltr.0806.06

Clark, T., Spiby, R., & Tasviri, R. (2021). Crisis, collaboration, recovery: IELTS and COVID-19. *Language Assessment Quarterly*, *18*(1), 17-25. https://doi.org/10.1080/15434303.2020.1866575

Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative*. Pearson Education Limited.

Crocetta, C., Antonucci, L., Cataldo, R., Galasso, R., Grassia, M. G., Lauro, C. N., & Marino, M. (2021). Higher-order PLS-PM approach for different types of constructs. *Social Indicators Research*, *154*, 725-754. https://doi.org/10.1007/s11205-020-02563-w

Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, *45*, 62-69. https://doi.org/10.1016/j.stueduc.2015.03.009

Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, *64*, 100824. https://doi.org/10.1016/j.stueduc.2019.100824

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation*. Routledge.

Edgar, T., Manz, D., & Manz, D. (2017). Exploratory study. *Research methods for cyber security*, *29*, 95-130.

Ehtsham, M., Zeb, A., Alam, A., & Ahmad, S. (2023). Examining the Impact of Socioeconomic Factors on English Language Proficiency: A Cross-Cultural Analysis. *Al-Mahdi Research Journal (MRJ)*, *5*(2), 126-140.

Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre‑sessional language courses. *Assessment in Education*, *14*(1), 75-97. https://doi.org/10.1080/09695940701272880

Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, *13*(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Guinot, C., Latreille, J., & Tenenhaus, M. (2001). PLS path modelling and multiple table analysis. Application to the cosmetic habits of women in Ile-de-France. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 247-259. https://doi.org/10.1016/S0169-7439(01)00163-0

Ha, N. T. T. (2019). A literature review of washback effects of assessment on language learning. *Ho Chi Minh City Open University Journal of Science - Social Sciences*, *9*(2), 3-16. https://doi.org/10.46223/HCMCOUJS.soci.en.9.2.257.2019

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning*, *46*(1-2), 1-12.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, *31*(1), 2-24. https://doi.org/10.1108/EBR-11-2018-0203

Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language Testing*, *14*(3), 295-303. https://doi.org/10.1177/026553229701400306

Heck, R., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches*. Routledge.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1108/F-11-2014-0081

Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, *6*(2), 151-167. https://doi.org/10.1080/15434300902801925

Hughes, A. (1993). Backwash and TOEFL 2000. *Unpublished manuscript, University of Reading*.

Hughes, A. (2002). *Testing for Language Teachers* (2 ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511732980

Hung, S.-T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, *25*(1), 21-36. https://doi.org/10.1080/09588221.2010.551756

IELTS. (2021). *New milestone for world-leading English test: IELTS trusted by more than 11,000 organisations across the globe*. Retrieved 10 May 2024 from https://www.ielts.org/news/2021/ielts-trusted-by-more-than-11000-organisations-across-the-globe

Johnson, D. D. (2008). *Stop high-stakes testing: An appeal to America's conscience*. Rowman & Littlefield.

Law, K. S., Wong, C.-S., & Mobley, W. M. (1998). Toward a taxonomy of multidimensional constructs. *Academy of management review*, *23*(4), 741-755. https://doi.org/10.5465/amr.1998.1255636

Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, *68*, 100967. https://doi.org/https://doi.org/10.1016/j.stueduc.2020.100967

Madaus, G. F., & Keillor, G. (1988). The influence of testing on the curriculum. *Teachers college record*, *89*(5), 83-121. https://doi.org/10.1177/016146818808900505

Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet? *MIS*

quarterly, iii-ix. https://doi.org/10.2307/25148727

Mesly, O. (2015). Simulation and Quantitative Techniques. In O. Mesly (Ed.), *Creating Models in Psychological Research* (pp. 53-61). Springer International Publishing. https://doi.org/10.1007/978-3-319-15753-5_7

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256. https://doi.org/10.1177/026553229601300302

Nguyen, H., & Vu, T. (2020). *Phân tích dữ liệu áp dụng mô hình PLS-SEM [Data analysis using PLS-SEM model]*. NXB Kinh Tế Thành Phố Hồ Chí Minh.

Nguyen, H. N., & Nguyen, D. K. (2022). Vietnamese Learners' Performance in The IELTS Writing Task 2: Problems, Causes, and Suggestions. *International Journal of TESOL & Education*, *2*(1), 170-189. https://doi.org/10.54855/ijte.222111

Nguyen, H. T. M. (2023). The Washback of the International English Language Testing System (IELTS) as an English Language Proficiency Exit Test on the Learning of Final-Year English Majors. *TESL-EJ*, *27*(2), n2. https://doi.org/10.55593/ej.26106a8

Nhan, T. (2013). The TOEIC® test as an exit requirement in universities and colleges in Danang City, Vietnam: Challenges and impacts. *International Journal of Innovative Interdisciplinary Research*, *2*(1), 33-50.

Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of ielts preparation materials from a critical perspective. *Language Assessment Quarterly*, *18*(4), 382-397. https://doi.org/10.1080/15434303.2021.1883618

Pan, Y., & Newfields, T. (2011). Teacher and student washback on test preparation evidenced from Taiwan's English certification exit requirements. *International Journal of Pedagogies and Learning*, *6*(3), 260-272. https://doi.org/10.5172/ijpl.2011.6.3.260

Pan, Y.-C., & Newfields, T. (2012). Tertiary EFL Proficiency Graduation Requirements in Taiwan: A Study of Washback on Learning. *Electronic Journal of Foreign Language Teaching*, *9*(1), 108–122. https://doi.org/10.5172/ijpl.2011.6.3.260

Pearson, W. S. (2019). Critical perspectives on the IELTS test. *ELT Journal*, *73*(2), 197-206. https://doi.org/10.1093/elt/ccz006

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European journal of education studies*, *3*(9), 369-387.

Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. In *International English Language Testing System (IELTS) Research Reports 2003: Volume 4* (pp. 153-191). IDP: IELTS Australia Canberra.

Ren, Y. (2011). A study of the washback effects of the College English Test (band 4) on teaching and learning English at tertiary level in China. *International Journal of Pedagogies and Learning*, *6*(3), 243-259. https://doi.org/10.5172/ijpl.2011.6.3.243

Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs: Tutorial and annotated example. *ACM sigmis database: The database for Advances in Information Systems*, *40*(3), 9-39. https://doi.org/10.1145/1592401.1592405

Roopa, S., & Rani, M. S. (2012). Questionnaire designing for a survey. *Journal of Indian Orthodontic Society*, *46*(4), 273-277. https://doi.org/10.5005/jp-journals-10021-1104

Sadler, D. R. (2016). Three in-course assessment reforms to improve higher education

learning outcomes. *Assessment & Evaluation in Higher Education*, *41*(7), 1081-1099. https://doi.org/10.1080/02602938.2015.1064858

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of market research*, *26*(1), 1-40. https://doi.org/10.1007/978-3-319-57413-4_15

Schunk, D. H. (2012). *Learning theories an educational perspective*. Pearson Education, Inc.

Shamsuddin, K., Fadzil, F., Ismail, W. S., Shah, S. A., Omar, K., Muhammad, N. A., Jaffar, A., Ismail, A., & Mahadevan, R. (2013). Correlates of depression, anxiety and stress among Malaysian university students. *Asian J Psychiatr*, *6*(4), 318-323. https://doi.org/10.1016/j.ajp.2013.01.014

Sharma, H., & Aggarwal, A. G. (2019). Finding determinants of e-commerce success: a PLS-SEM approach. *Journal of Advances in Management Research*, *16*(4), 453-471. https://doi.org/10.1108/JAMR-08-2018-0074

Shih, C.-M. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, *64*(1), 135-161. https://doi.org/10.3138/cmlr.64.1.135

Tae-Young, K., & Yoon-Kyoung, K. (2016). A quasi-longitudinal study on English learning motivation and attitudes: The case of South Korean students. *Journal of Asia TEFL*, *13*(2), 72-166. https://doi.org/10.18823/asiatefl.2016.13.2.5.138

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational statistics & data analysis*, *48*(1), 159-205. https://doi.org/10.1016/j.csda.2004.03.005

ThanhNienNews. (2022). *"Cơn sốt" cho con đi học IELTS*. Retrieved 10 May 2024 from https://thanhnien.vn/con-sot-cho-con-di-hoc-ielts-1851506468.htm

Tran, Q. H. (2021). An investigation into non-English major students' problems in taking Aptis listening and reading. *International Journal of TESOL & Education*, *1*(3), 176-191. https://doi.org/10.11250/ijte.01.03.010

Tsagari, D., & Cheng, L. (2017). Washback, Impact, and Consequences Revisited. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language Testing and Assessment* (pp. 359-372). Springer International Publishing. https://doi.org/10.1007/978-3-319-02261-1_24

Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing, 39*(2), 212-238. https://doi.org/10.1177/02655322211050600

Wall, D. (2012). Washback in Language Assessment. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal1274

Watanabe, Y. (2004). Methodology in washback studies. In *Washback in language testing* (pp. 41-58). Routledge.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS quarterly*, *33*(1), 177-195.

Xiao, W. (2014). The Intensity and Direction of CET Washback on Chinese College Students' Test-taking Strategy Use. *Theory & Practice in Language Studies (TPLS)*, *4*(6), 1171-1177. https://doi.org/10.4304/tpls.4.6.1171-1177

Xiaoju, L. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual & Multicultural Development*, *11*(5), 393-404. https://doi.org/10.1080/01434632.1990.9994425

Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, *50*, 56-68. https://doi.org/10.1016/j.system.2015.03.002

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, *30*(1), 49-70. https://doi.org/10.1177/0265532212442634

Xu, Q., & Liu, J. (2018). Washback: Definitions and Dimentions. In Q. Xu & J. Liu (Eds.), *A Study on the Washback Effects of the Test for English Majors (TEM): Implications for Testing and Teaching Reforms* (pp. 17-22). Springer Singapore. https://doi.org/10.1007/978-981-13-1963-1_2

Zhan, Y., & Wan, Z. H. (2016). Test Takers' Beliefs and Experiences of a High-stakes Computer-based English Listening and Speaking Test. *RELC Journal*, *47*(3), 363-376. https://doi.org/10.1177/0033688216631174

Zhengdong, G. (2009). IELTS Preparation Course and Student IELTS Performance:A Case Study in Hong Kong. *RELC Journal*, *40*(1), 23-41. https://doi.org/10.1177/0033688208101449

Biodata

Nguyen Thanh Minh is a lecturer at Van Lang University, Ho Chi Minh, Vietnam. He has nearly ten years of experience teaching English to both English majors and non-English majors. He is also pursuing his PhD in TESOL at HCMC Open University. His research interests are curriculum design, assessment, predictive modeling, and professional development.

EFL postgraduate students' perceptions on the use of Grammarly and peer feedback to improve their academic writing skills

Ha Yen Nhi[1*], Ho Ngoc Phuong[2]

[1] Language Institute, Van Lang University, Ho Chi Minh City, Vietnam
[2] Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam
[*] Corresponding author's email: nhi.hy@vlu.edu.vn
[*] https://orcid.org/0000-0002-5554-2937
https://doi.org/10.54855/ijte.25512

ABSTRACT

Keywords: EFL postgraduate students, perception, Grammarly, peer feedback, peer response activities, academic writing skills

Grammarly and peer feedback have recently become two evaluative approaches that are commonly used in writing classes to provide effective comments on students' writing (Fahmi & Cahyono, 2021). However, recent studies have only examined their effectiveness on students' writing skills, neglecting their thoughts and perceptions. To address this gap, the paper explores EFL postgraduate students' perceptions of using Grammarly and peer feedback activities to enhance their academic writing skills. The interview approach was incorporated to collect data, using the participation of 10 EFL postgraduate students who were learning at Van Lang University. The qualitative study indicates that students feel satisfied with both approaches; however, they all claimed that although their peers take more time to complete revision, those comments are more in-depth and constructive. Grammarly is fast but sometimes inappropriate and limited. Moreover, more learners prefer peer feedback. It is highly recommended that this study serve as a database for further quantitative research on other groups of participants.

Introduction

In this day and age, the teaching of writing has dramatically changed its focus from writing outcomes to the writing process. Hence, providing feedback has become a crucial aspect for EFL learners. In the past, educators were responsible for offering comments on their students' writing tasks. However, the introduction of peer feedback has emerged as a new approach to further develop students' writing skills (Asper et al., 2024; Zeevy-Solovey, 2024). Generally, peer feedback is an evaluative process where students receive assessments and grades from their peers (Falchikov, 2001; Pham et al., 2020) with this type of activity, students are required to make comments on their friends' works in written or oral form with the purpose of improving them (Liu & Hansen, 2002). Moreover, several researchers have proven that this type of evaluation can bring a whole host of advantages to L2 classes, especially the writing ones. First of all, one of the major benefits of applying peer response activities in writing classes is that

students can effectively fix their writing thanks to comments made by their peers (Robinson, 2005; Wakabayashi, 2013; Pham et al., 2020; Latifi et al., 2023). Along the same line with that, Pratama and Arriyani (2021) found that even students with low motivation in studying are still able to enhance their writing abilities through the implementation of peer response activities. Besides that, Liu and Carless (2006) argue that this kind of assessment can be beneficial in developing students' detection and revising their work.

Since the emergence of technological advancements, the Automated writing evaluation (AWE) program, a computer-based platform, has been the center of attention. In general, this type of program has the ability to utilize sophisticated language analysis methods to offer writers immediate, detailed, and comprehensive feedback on their lexicon, grammar, and spelling, facilitating improvements in their writing (Warschauer & Ware, 2006; Cotos, 2011; Grimes & Warschauer, 2010). Moreover, by using a technique called "natural language processing", some programs recently are even able to process and diagnose an overall score for people's writing works (Shermis & Burstein, 2003, p37; Grimes & Warschauer, 2010; Hockly, 2019). Thanks to those feedback and scoring mechanisms, students can save time in previewing their works and then have suitable corrections, leading to enhanced versions of their writing (Stevenson & Phakiti, 2014; Parra & Calero, 2019; Fahmi & Cahyono, 2021). Overall, the AWE program is a perfect assistant for EFL learners in terms of analyzing, evaluating, and scoring.

Among the different Automated Writing Evaluation (AWE) programs available, Grammarly stands out as a highly commendable choice due to its valuable features and user-friendly interface, making it a recommended tool for integrating into writing classes (Fahmi & Cahyono, 2021). The founders of Grammarly (2020) emphasize its reputation as a user-friendly and effective learning aid that assists learners in addressing their writing needs. For better clarity and readability, Grammarly can provide insightful feedback that generates corrections and suggestions for better clarity and readability by proficiently detecting all kinds of errors (spelling, grammar, and punctuation). The platform is renowned for its exceptional accuracy in evaluating written content, ensuring error-free and impactful writing.

Statement of the problem

In general, peer feedback or Grammarly feedback offers notable advantages in different sectors, and both methods significantly contribute to the improvement of writing skills and outcomes among EFL learners. However, the study conducted by Fahmi and Cahyono (2021) reveals that students express partial satisfaction with the use of Grammarly feedback alone; instead, they prefer a combination of feedback from their teachers and Grammarly. Furthermore, Ghufon (2019) states the application of the Grammarly platform in EFL writing has been shown to have a positive impact on error reduction, but the website is not as effective in detecting the content of students' writing, a task that peers are capable of finishing (Pham & Usaha, 2016).

On the other hand, cultural factors pose a significant challenge to the effectiveness of peer feedback (Chareonsuk, 2011). In Asian countries, where relationships are built on mutual respect, individuals tend to avoid actions that may cause others to lose face (Chareonsuk, 2011); hence, students often feel hesitant to provide comments when being asked to evaluate their peers' writing; resulting in comments that lack quality and sincerity (Kunwongse, 2013). Moreover, the lack of guidelines can lead students to focus on surface errors rather than content, potentially hindering meaningful revisions and improvements. Another issue is that since providing feedback on students' work has traditionally been a teacher's responsibility, students often lack the necessary assessment skills and feel reluctant to evaluate their peers' work when asked to do so (Le, 2023). Last but not least, younger and less experienced students may struggle to provide constructive feedback, leading to ineffective learning experiences (Hutt et al., 2024).

The purpose of the study

With all of the problems mentioned above, the authors may examine how EFL postgraduate students feel and think about peer response activities and Grammarly feedback and in which aspects they are satisfied or dissatisfied with the two methods. Ultimately, the authors aim to identify which approach provides more constructive feedback for students' writing based on their opinions. Therefore, the paper's primary objective is to analyze EFL postgraduate students' perceptions of the implementation of Grammarly and peer feedback to enhance their writing skills.

The significance of the study

This paper serves as a scientific record that represents the perspectives of EFL postgraduate students on utilizing Grammarly feedback and peer feedback to enhance their writing abilities. In addition, this research aims to explore the incorporation of blended feedback in an academic writing class. Hence, this study hopes to make a great contribution to the current body of knowledge regarding the use of Grammarly feedback and peer response activities in various settings and subjects. Moreover, it suggests the potential for future researchers to integrate this study into their own investigations to attain a more comprehensive understanding within the same field.

Literature review

Definition of "academic writing skills"

According to Jones (1994, as cited in Iftanti, 2016), writing skills are defined as (1) the ability to choose appropriate words and grammatical structures for different purposes and topics; (2) the ability to state clear ideas and organize them in order while still maintain the coherence between sentences, paragraphs and parts of articles; (3) the ability to correct writing errors.

In a broad context, writing can be classified into two distinct branches: academic writing and creative writing. Academic writing is distinguished from creative writing to some extent. Specifically, while the latter focuses mostly on creativity and the use of informal language (slang or abbreviations), the former is mainly about the structured organization of sentences and formal language (Oshima & Hogue, 2007, p3). As stated by Irvin (2010), good academic writing is one that can necessitate the demonstration of comprehensive knowledge and the display of adeptness in specific cognitive abilities such as critical thinking, interpretation, and proficient presentation within the context of disciplinary domains.

Definition of "perception"

Previous papers have claimed that there is no definition of "perception" and that it may vary from researcher to researcher. Efron (1969) found that perception serves as the fundamental way for individuals to mentally connect with the world surrounding them, and all our conceptual understanding is built upon and originates from this initial mode of consciousness. In the same line, Nurzakiah (2021) finds that perception roots in "precipice" - a Latin word that involves receiving input and interpreting signals to have experiences or make connections.

Crane (2005) argues that perception is the thoughts of people about the world around them formed by the five senses (hearing, listening, touching, tasting, and watching). Similarly, in the paper of Epstein et al. (2023), perception in humans refers to the transformation of sensory input into structured and meaningful experiences that result from the cooperation of sensory stimulation and the underlying cognitive processes. In addition, according to the Oxford

Learner's Dictionaries, perception is defined in different ways: (1) an image or a belief that is formed as a result of an individual's understanding of something; (2) a way people think or feel about something, especially through five senses.

According to Ghadirian et al. (2018), perception is characterized as a series of actions with the purpose of acquiring knowledge or information within the field of education. This process can occur through exposure to different environments experienced by learners. Freiberg (1999) even emphasizes that learners' perception is a primary element of education improvement.

Despite the various definitions surrounding perception, this study emphasizes perception as a subjective thought of EFL postgraduate learners influenced by a long period of experience with two evaluative methods.

Local and global revision

It is a fact that there are two distinct areas in writing, global and local, and several differences are pointed out between them. About the definitions, on the one hand, people engage in local revision when they just change one to two words in a sentence, which can just have an impact only on a few sentences, while global revision involves making changes to one part of a passage that subsequently requires modifications in other parts of it (Ramage et al., 2011). In the same line, local revision refers to the connections between neighboring clauses within a text, but global cohesion refers to the cohesive elements and structures that connect and unify larger discourse units such as paragraphs and sections (Van Dijk & Kintsch, 1983).

In terms of areas of focus, the main focus of global revision includes ideas, purpose, cohesion, content, reader, coherence, and organization (Ramage et al., 2011). Similarly, Bridwell (1980) supported that global revisions encompass substantial alterations carried out at different levels, including individual sentences, paragraphs, and the text's overall structure. These revisions primarily focus on the meaning and semantics of the text and require actions such as removing, relocating, adding, changing, or fixing ideas across the text. On the other hand, local revision focuses on addressing errors related to spelling, grammar, and mechanics (Ramage et al., 2011). From the view of Hayes (2000), the researcher considers the process of local revision as the problem-solving activity in which reviewers skim through a writing outcome to detect and fix mistakes and may ignore the comprehension factor of the writing.

Regarding the importance of local and global revision, it is argued by Bransford and Johnson (1972, as cited in McNamara et al., 2002) that texts that are just locally revised can hinder understanding, texts that maintain a clear overall structure but lack coherence due to grammatical or spelling errors can pose challenges for reading and comprehension. As opposed to the previous finding, Crossley and McNamara (2011) argue that just global revision can benefit the quality of writing. In the same way, Butler and Britt (2011) also claimed that the effectiveness of local revision in enhancing the writing quality of students' works is not comparable to that of global revision.

Based on those previous definitions and arguments about local and global revision, the research will analyze EFL postgraduate students' perceptions of local and global revision comments provided by Grammarly feedback and peer feedback to determine which element is more effective and preferable.

Previous studies

Numerous studies conducted around the world have explored how EFL learners perceive the utilization of Grammar feedback and peer response activities to develop the quality of their writing achievements.

The perception of learners toward peer feedback

Regarding peer feedback, through semi-structured interviews, Ebadi and Rahimi (2017) found that their participants feel positive about the implication of traditional and online peer editing activities in their writing classes. It has been demonstrated that both approaches contribute positively to enhancing students' academic writing capabilities. In the same vein, Huisman and co-authors (2018) claim that regardless of whether they are receiving or giving feedback, participants consistently reported that such comments positively develop the quality of their writing. In addition, participants expressed a preference for explanatory feedback over analytical one.

Although it is a big concern that students who lack prior exposure to peer feedback can encounter several problems and think negatively about the method, Kasch et al. (2021) prove the opposite view. According to their study, even students who have never encountered peer feedback before are willing to use peer feedback and acknowledge the method's usefulness. In the context of online learning, peer feedback still expresses its potential. The findings of Noroozi and Mulder (2017) and Taghizadeh Kerman et al. (2022) have demonstrated a noteworthy correlation between student's perception of the value and credibility of peer feedback and their overall satisfaction with their learning experience. Similarly, in the study about integrating blog-based peer comments to improve learners' writing skills conducted in Vietnam, Pham and Usaha (2016) found that EFL students treasure feedback from their classmates. As explained by Misiejuk et al. (2021), students who find peer feedback to be valuable are more inclined to accept it, and this acceptance is evident through their acknowledgment of mistakes, expressing a wish to make revisions, recognizing the effectiveness of the feedback provided by their peers. More than developing learners' writing competence, students themselves can even trigger their review and critique abilities, which is a foundation for critical thinking (Lee, 2017). Overall, it is determined that students at different levels perceive the activity positively and recognize the potential of the method to enhance their writing abilities.

However, it is unavoidable that some students may doubt the accuracy and validity of feedback from their colleagues. As found by Taghizadeh Kerman et al. (2022) and Burgess et al. (2013), several participants claimed that they lacked belief and confidence in their peers' knowledge and ability, which made them express their hesitancy to engage in peer feedback practices. Hence, students gravitate toward expert feedback (Tai et al., 2015). Another reason leading to negative perceptions is that due to the lack of experience in rhetoric, learners tend to prioritize sentence-level aspects over ideas and organization of the writing, resulting in feedback that may be less helpful in improving overall writing quality (Hyland, 2003). Lastly, although constructive criticism is beneficial for students' writing, they have the tendency to ignore this type of feedback, as they do not want to feel negative (Ryan & Henderson, 2017).

Ho and his colleagues (2020) have researched the viewpoints of Vietnamese lecturers and their learners regarding written peer feedback in Vietnam. From the findings, the researchers point out that even EFL students who have never experienced peer feedback activities before still perceive it positively, as they acknowledge its value in providing opportunities for learning from peers and improving their writing competence. It has also been found that apart from improving students' writing abilities, peer feedback is useful for both receivers and givers (Dang, 2024). Moreover, Dang (2024) believes that through collaboration in peer feedback, users can develop their critical thinking skills, improve their social interactions, and strengthen their intellectual reasoning, which is necessary for their future careers.

On the contrary, Pham (2020) argues that even though peer editing activities have obtained

encouragement from EFL learners, some students still prefer feedback from instructors because they feel hesitant and uncertain about giving feedback on their peers' work. In the same vein, Vo (2022) claims that the preference for teacher correction also comes from the belief in teachers' academic abilities and the ability to deliver precise and detailed corrective feedback to students. Another issue Pham (2020) highlighted regarding the lack of enthusiasm for peer feedback activities is students' concern about embarrassing their peers. As a result, they are unwilling to join peer response activities.

The perception of learners towards Grammarly feedback

In terms of Grammarly feedback, most recent studies reveal that Grammarly has received a whole host of positive responses from users, particularly in writing classes. According to Fahmi and Cahyono's (2021) study, the participants shared their views on utilizing Grammarly to develop their writing proficiency; the results reveal that real-time feedback provided by Grammarly greatly influences students' writing development and helps save their time. In the same way, participants in the study conducted by Huang and co-researchers (2020, May) claim to like using Grammarly to revise their writing during writing classes; furthermore, their knowledge about grammatical points and writing structure also develops. In the same vein, Ghufron and Rosyida (2018) and Ghufon (2019) prove that with the incorporation of Grammarly, the number of spelling, grammar, and punctuation errors has decreased significantly.

Grammarly has proven its role as a useful evaluating assistant for EFL learners and a great assistant for English teachers. According to Wilson and Andrada (2016) and Lailika (2019), because Grammarly helps to revise students' writing works, teachers are able to noticeably reduce their workload and allocate more time towards supporting learners with their writing structure and organization.

However, it is pointed out that there are several concerns with Grammarly feedback in terms of validity and correctness. Some EFL students in the research of O'Neill and Russell (2019) expressed their negative feelings towards using Grammarly in writing tasks; the matter is that when the students used passive voice, some complex structures, or several word choices, Grammarly recommended them to modify for stylistic reasons rather than for accuracy, and the platform even failed to detect some errors. For example, sometimes students use passive voice structures, but Grammarly still suggests they convert to active ones, which may overshadow their unique voice and style. Besides that, students occasionally could not understand Grammarly's suggestions. Furthermore, Lailika (2019) and Fahmi and Cahyono (2021) reveal that learners who were not good at English expressed dissatisfaction with Grammarly, as those comments are misleading, which may cause difficulties for them. Another concern about this app is its limitations in accessing content and organization. According to Ghufron and Rosyida (2018), the website is not very effective when it comes to organizing content. This is because the system cannot tell if the student's writing stays on a topic or if their ideas are arranged in a logical way. Finally, many authors agree that the original version of Grammarly is not as effective as the premium one due to some limitations, but purchasing this version seems to be a financial burden for some students (Fitria, 2021; Fitriana & Laeli, 2022; Dewi, 2023).

Research gaps

Overall, it is clear that several gaps can be observed from previous studies. First, existing papers predominantly focus on examining students' perceptions of peer feedback (Gaynor, 2020; Misiejuk et al., 2021) or Grammarly feedback separately (O'Neill & Russell, 2019; Fitriana & Nurazni, 2022). In the end, there is a lack of materials investigating the perceptions of learners

on the combination of these two techniques. Next, most of the existing research has chosen undergraduate students and students at lower proficiency levels as their primary participants (Pham et al., 2020; Aidil, 2021; Fitriana & Nurazni, 2022), but there are nearly no research records about the utilization of Grammarly feedback and peer feedback, particularly at the postgraduate level.

Although there are numerous papers related to this field worldwide, not many of them are conducted in Vietnam, especially those about students' perceptions of Grammarly feedback. Additionally, Vietnamese researchers predominantly concentrate on investigating the usefulness or the pros and cons of such approaches, but not the perception. Consequently, it is urgent to conduct this study.

Research questions

1. What are the perceptions of EFL postgraduate learners about feedback from peers to enhance their writing quality?
2. What are the perceptions of EFL postgraduate learners about feedback from Grammarly to enhance their writing quality?
3. Which method is more effective for EFL postgraduate students, between Grammarly and peer feedback?

Methods

Participants

The study took place during the third semester of the academic year 2022-2023, within the Research Writing class at the Faculty of Foreign Languages of a prestigious university in Ho Chi Minh City, Vietnam—Van Lang University.

The sampling technique that was applied was the convenience sampling technique. According to Andrade (2021), convenience sampling is a non-probability sampling technique where samples are drawn from a population that is easily accessible to the researcher. In other words, convenience sampling involves selecting readily available participants, such as students in a classroom or individuals in a specific location. The researchers selected this sampling method because their classmates were readily accessible, facilitating observation and data collection. Additionally, conducting the study with Master's students from other classes was not feasible due to scheduling conflicts.

The class consisted of 10 students, all of whom held a Bachelor's degree in language-related fields and are currently pursuing a Master's degree program. The age range of the participants ranged from 23 to 32 years old. Furthermore, these EFL postgraduate students have accumulated over 10 years of experience in English language learning. In terms of the participants' English proficiency, it is worth noting that each learner had to meet a minimum requirement of B2 CEFR level or higher in English to be accepted into the Master's program, as followed by the entrance requirements. It is worth noting that their primary motivations for learning English included job promotion and further academic pursuits.

Design of the study

A qualitative study was conducted with 10 EFL postgraduate students' participation in a Research Writing class. During the Research Writing class, the learners were taught how to write various types of paragraphs and essays in an academic and professional way. Following each lecture, the EFL postgraduate students were assigned a writing task that they would

collaboratively complete in groups, as required by the lecturer. When completed, the assignments would be reviewed by both Grammarly and the other colleagues in the class. With the feedback received, the students then revised their work before submitting it to the lecturer.

On the final day of the course, the students participated in structured interviews with the researchers so as to deepen their perceptions of these two methods. The reason why we chose to utilize this interview method was due to its advantages. According to Lune and Berg (2017), this type of interview allows researchers to effectively gather information in terms of students' thoughts and attitudes on study-related issues. Additionally, Peus et al. (2013) argued that the structured approach provides a specific context for interviewees, which facilitates a more customized evaluation. Last but not least, as compared to unstructured interviews, structured ones offer a higher level of validity and reduce the potential for additional risks (Levashina et al., 2014).

Procedure

The study was conducted at the beginning of the third semester of the academic year 2022-2023. Prior to conducting the research, the author obtained permission from Prof. Vu Phi Ho Pham, the lecturer responsible for teaching the Research Writing course. At the first session of the course, the lecturer provided an overview of the course syllabus, introduced the notion of peer feedback, as well as guided the students on effectively integrating peer feedback into their writing process. After that, the researchers provided the students with comprehensive instructions on the utilization of Grammarly as a tool for assessing and reviewing their assignments.

After the course introduction, 10 EFL postgraduate students participating in the study were put into four groups to facilitate peer feedback activities. Two groups consisted of three members each, while the remaining two groups comprised two members each. The purpose of grouping them is for peer feedback activities afterwards, and it is crucial to note that the groups were carefully formed to ensure that the members possessed similar language proficiency levels. All participants achieved a good to excellent Bachelor's degree from different universities across Vietnam.

Students would be assigned a group writing task after each lesson to collect data for the research. For the first two weeks, they would write opinion paragraphs to discuss the following topics: "Collaborative learning" and "No one is perfect". In the 3rd week, an argument essay about "The implementation of ChatGPT in learning" was given to EFL postgraduate students. In the following week, the participants were required to write a cause-effect paragraph about the topic "The effects of technology or mobile devices on L2 learning". During the 5th and 6th sessions, the EFL postgraduate learners were tasked with completing summary and critique paragraphs based on a paper provided by the lecturer. The students were expected to write their final paper's introduction and literature review in the next two weeks. The final three weeks were for knowledge revision and teacher feedback.

Throughout the course, students completed eight writing assignments. Before submitting their work to the lecturer, each group was encouraged to seek feedback from two other groups. The EFL postgraduate students used a checklist when reviewing their peers' work to ensure that the feedback was constructive and helpful. When giving feedback on their peers' writing, the students were asked to focus on four elements: task achievement, coherence and cohesion, grammatical structures, and lexical resources. For the task achievement criterion, students should evaluate how effectively the writer responds to the prompt, develops ideas, and includes relevant examples or arguments. Secondly, in terms of coherence and cohesion, students need

to analyze the logical arrangement of ideas, the flow between sentences and paragraphs, and the appropriate use of linking words to create cohesive and well-structured writing. Furthermore, the grammatical structures criterion emphasizes the diversity and accuracy of sentence forms, ranging from simple to complex, ensuring that any errors do not hinder understanding. Lastly, students assess the writer's ability to employ a wide range of vocabulary accurately, including less common vocabulary, while maintaining clarity and avoiding repetition or misuse of words.

At the same time, students uploaded their assignments to Grammarly for further editing. This feedback and revision process took place outside of class and before the submission of the final homework assignments. Once feedback from both peer evaluations and Grammarly was incorporated, students revised their work and submitted the edited versions. Additionally, they were asked to record their feelings and opinions about the process to support later interviews.

After experiencing peer feedback and Grammarly feedback for the whole course, the researchers conducted structured interviews with each participant on the final day of the course. The aim was to explore their opinions and feelings about using these two methods to assess their writing. The interviews were conducted in a face-to-face setting within the classroom and lasted approximately 20 to 30 minutes. The researcher needed to obtain permission from the 10 EFL postgraduate students to conduct interviews.

Data collection and analysis

The figures were gathered through structured interviews. At the first stage of each interview session, the author spent a few minutes breaking the ice and collecting some of their personal information, including their age, current job, and duration of English language learning. When ensuring that the interviewees were completely comfortable, the author provided an overview of the study, including the title, purpose, and other relevant details. The author also encouraged the interviewees to provide honest responses from what they had experienced so far so as to avoid bias. The author also used a phone to record each interview section to facilitate the data collection process, which was also informed to the participants.

Overall, the authors interviewed the participants with a total of 16 questions divided into three main sections. Before coming to the final questions list, the researchers piloted it many times with support and feedback from Dr. Vu Phi Ho Pham. The first section, comprising seven questions, aimed to gather the participants' perceptions of Grammarly feedback. The second part also consisted of 7 questions about students' perception of peer feedback. In the first two sections, there are five open-ended questions, one yes-no question, and one question in the form of the five-point Likert scale, including Very satisfied, Satisfied, Neutral, Dissatisfied, and Very Dissatisfied to collect data about the level of satisfaction of the participants towards the implementation of Grammarly and peer feedback in enhancing their writing competence. In the final section, the interviewees needed to choose the preferred method and explain why. These interviews wished to comprehensively explore students' perceptions, so each interview lasted about 20 to 30 minutes.

After having all of the data, the author started to listen to the recorded audio and noted down the answers in a Word document for analysis.

Results/ Findings

This section will show readers data resulting from interviews. The data analysis was conducted using the Statistical Package for Social and Science (SPSS) version 22, and the findings were visualized through tables and pie charts. The analysis includes numerical measurements such as mean, percentage, and St.derivation (SD). Five headings, like demographic information, focused areas of Grammarly and peer feedback, peer feedback, Grammarly feedback, and peer feedback vs. Grammarly feedback, clarify the findings.

Demographic information

Table 1

The personal information of participants

Student	Gender	Group
Student 1	Female	Group 1
Student 2	Female	
Student 3	Female	
Student 4	Female	Group 2
Student 5	Male	
Student 6	Female	Group 3
Student 7	Female	
Student 8	Female	
Student 9	Female	Group 4
Student 10	Male	

As stated above, 10 EFL postgraduate students participated in the interviews. Most of them were female, while there were only two male students.

Focused areas of Grammarly feedback and peer feedback

Charts 1 and 2

Focused areas of feedback
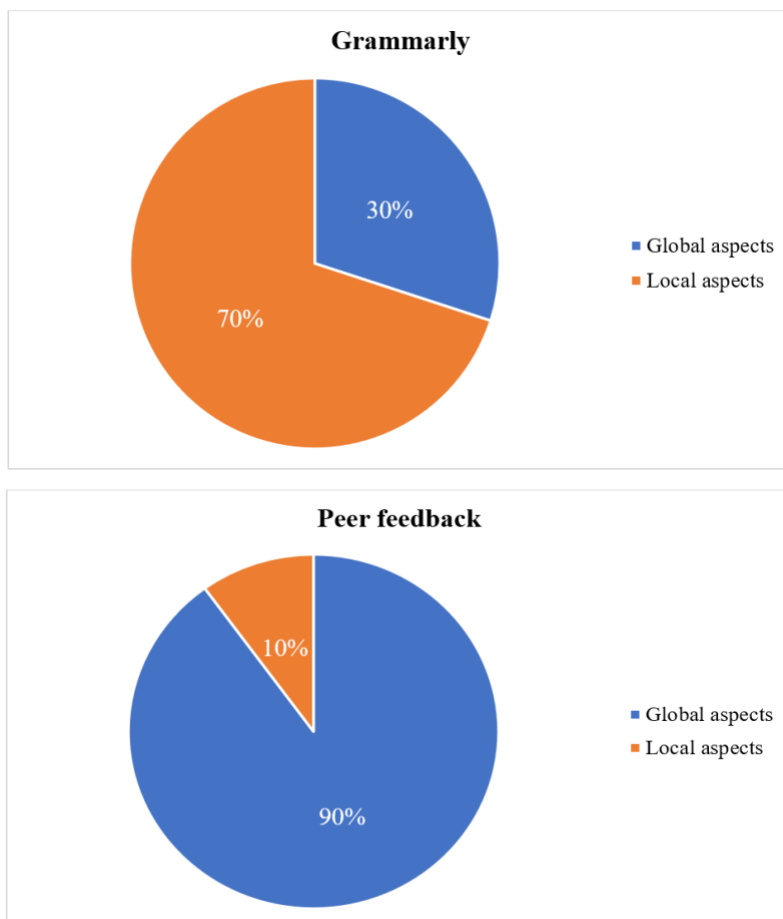
The first two pie charts illustrate the focused areas of feedback that EFL postgraduate students obtained from both Grammarly and peer feedback. Based on the course curriculum, there are 13 factors of writing assessment, including Grammar, Vocabulary, Word order, Punctuation, Capitalization, Spelling, Collocation, Ideas, Content, Cohesion, Organization, and Citation. These factors are categorized into two main groups: Global aspects and Local aspects. The global aspects include Ideas, Content, Organization, Cohesion, and Citation. The remaining factors belong to local aspects.

According to the interviews, EFL postgraduate students reported that the majority of feedback they received from Grammarly was related to local revision (70%), and only 30% of those comments were global revision. On the other hand, whereas their classmates concentrated more on global revision, accounting for 90%, the remaining 10% of the feedback addressed local aspects. Overall, it can be seen that while Grammarly predominantly focused on local assessment, peer feedback primarily addressed global factors.

Peer feedback

Table 2.

Levels of satisfaction

Questions	VD%	D%	N%	S%	VS%	Mean	SD	Decision
How satisfied are you with peer feedback?	0	0	10	50	40	2	2.35	F

Table 2 illustrates data about the interviewees' satisfaction levels with peer editing activities. The mean score is 2, with most of the students expressing satisfaction with the method. Four students reported high levels of satisfaction, five students felt satisfied with peer feedback, and only one student had a neutral perspective. Fortunately, no students expressed negative feelings toward the peer response activities.

Why are you satisfied with peer feedback?

EFL postgraduate students expressed satisfaction with the feedback from their peers due to several beneficial aspects. First of all, thanks to high quality and constructive comments in terms of global factors from their classmates, their writing skills have improved in leaps and bounds. Students 8 and 9 specifically mentioned that the comments from their peers proved more practical and effective than the feedback from Grammarly. This was attributed to the peers' ability to consider the work's purpose, audience, and style, leading to suggestions that aligned better with the writers' intentions.

Secondly, the logic of the content is also very important. According to the responses of Student 6 and Student 9, their friends' global feedback guided them in choosing the best ideas for their writing works and organizing those ideas in a logical order.

The next factor is that the activities allowed them to broaden their knowledge and identify areas of weakness in their writing. Students 3, 4, 6, and 7 reported that they all expressed their interest in peer response activities, as they provided valuable opportunities for improvement and learning.

Next, Student 6 also emphasized that discussing the feedback obtained with peers enhanced their critical thinking abilities.

Last but not least, Student 1 emphasized the convenience of immediately meeting with classmates in class to discuss and gain further understanding when faced with comments that were initially unclear to them.

Have you ever encountered any difficulties with peer feedback? If yes, what are they?

The majority of EFL postgraduate students encountered several challenges with peer response activities. One of the primary problems was the feedback quality. The interviewees shared that certain peers did not approach the activities with seriousness, resulting in ambiguous or unhelpful comments, which did not improve the writing but made it worse. In addition, some of my classmates only gave normal and general praise, such as "good" or "excellent", without showing any errors in the writing. This led to dissatisfaction among the research participants, who expected more constructive feedback.

The second challenge identified was the limitation of time. Specifically, students were typically given approximately four days to leave comments on their peers' writing assignments, and this limited time often forced them to rush, leading to incomplete or low-quality evaluations. The students themselves also felt that the insufficient time did not allow them to provide thorough

and helpful comments.

Thirdly, although EFL postgraduate students in this Master's course were already good at English, varying levels of expertise and knowledge among peers could lead to confusing or unsuitable feedback. As a result, EFL postgraduate students needed to carefully select the ones that were most appropriate for their writing.

Lastly, peer feedback can also be influenced by biases, personal perspectives, subjective opinions, and cultural differences, which can result in feedback that does not align with the intended goals and objectives or the targeted audience.

Grammarly feedback

Table 3

Levels of satisfaction

Questions	VD%	D%	N%	S%	VS%	Mean	SD	Decision
How satisfied are you with Grammarly feedback?	0	10	30	60	0	2	2.55	F

Table 3 illustrates information about the participants' levels of satisfaction with Grammarly feedback. Overall, the mean score is 2, with the majority of EFL postgraduate students feeling satisfied with the implementation of Grammarly feedback. Three students felt neutral about the method, and only one student expressed the opposite idea.

Why are you satisfied with Grammarly's feedback?

In response to this question, Student 6 highlighted the convenience of Grammarly, emphasizing the simplicity of pasting their text and instantly receiving feedback within a minute, which is so fast. Students 3, 4, and 7 also recognized Grammarly's ability to identify grammar and spelling mistakes that their peers tended to overlook during self-editing. In the same vein, Student 10 even claimed that "because Grammarly follows predefined grammar rules and algorithms, offering a standardized evaluation, users can rely on its consistency and trust in its suggestions for error correction and language improvement ."Furthermore, due to real-time feedback, students can now save time and effort in editing and evaluating their writing works, as stated by Students 7 and 9.

Additionally, some students expressed the usefulness of Grammarly feedback in correcting punctuation errors, as the platform helped improve the clarity of their writing. Another aspect that pleased EFL postgraduate students was the app's ability to transform sentence structures to make the writing clearer, more powerful, engaging, and less wordy; for example, it would change passive voice structures into active voice. The overall predicted score was also a plus for the platform.

One of the last things the interviewees mentioned was Grammarly's accessibility and availability. The students explained that Grammarly is accessible online and through various platforms such as web browsers, desktop applications, and mobile apps, which enables users to receive assistance with their writing whenever they want.

Why are you dissatisfied with Grammarly's feedback?

Student 8 expressed her dissatisfaction with how Grammarly feedback was used in the Research Writing class for several reasons. First, the app seemed to focus more on correcting local aspects such as grammar, vocabulary, or spelling mistakes while ignoring global aspects like content, organization, cohesion, etc. This she personally did not appreciate since she believed that she

and her friends could handle them on their own.

Secondly, sometimes some of her sentences were grammatically wrong, but the website still claimed them as correct, causing confusion and frustration. This inconsistency undermined her trust in the accuracy of Grammarly's feedback.

The last reason was that some of the feedback from Grammarly resulted in changes that completely altered the meaning and intention of her original sentences.

Have you ever encountered any difficulties with Grammarly feedback? If yes, what are they?

9 out of 10 EFL postgraduate students claimed that they had encountered several challenges when using the Grammarly platform, and the most common one is about contextual understanding. Student 9 reported that Grammarly might sometimes struggle to understand the context or specific nuances of a particular sentence, leading to incorrect suggestions. Besides that, it primarily relies on patterns and rules, which may not accurately capture the intended meaning.

In the same vein as her, Student 10 claimed that Grammarly's primary focus is on grammar, spelling, and clarity, often overlooking the broader aspects of content, structure, and logical flow in a piece of writing. For example, Grammarly may not address issues related to the logical progression of ideas or provide suggestions for restructuring paragraphs to improve the flow of the text. As a result, Grammarly might suggest corrections that were not always appropriate for the particular context or writing style, as stated by Student 7.

In addition, when it comes to longer text pieces, the website was unable to detect structural and organizational issues, leading to frustration for Student 7.

Finally, citations are significant in academic writing, and the website's weakness is its inability to identify citation errors. As Student 6 reported, the website failed to highlight any citation mistakes that her peers easily pointed out.

Peer feedback vs Grammarly feedback

This part will provide information about which method is preferred by more EFL learners and the reasons for their choice.

Chart 3

A popular method among EFL postgraduate learners

Popular method among EFL postgraduate students

- Peer feedback 70%
- Grammarly feedback 30%
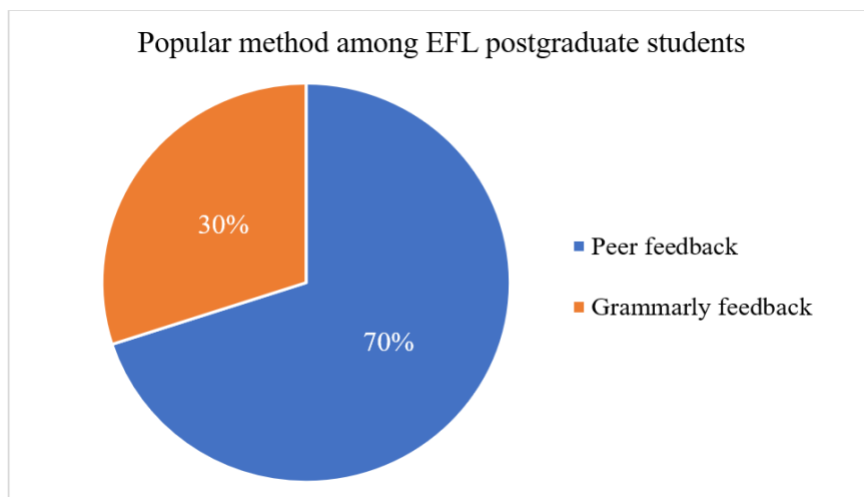
The pie chart above provides an insightful comparison of the preferences of EFL postgraduate learners regarding two different feedback methods, peer feedback and Grammarly feedback. The data highlights a clear trend in favor of peer response activities, with a substantial majority of the interviewees (70%) strongly prefer this method to develop their academic writing skills. On the other hand, a smaller proportion, comprising only 30% of the entire class, leaned towards using Grammarly feedback. Overall, it is clearly stated that peer feedback activities were more favorable towards postgraduate EFL learners than Grammarly.

Why did you choose peer feedback?

First and foremost, the reason why EFL postgraduate students prefer peer response activities was that they had a strong belief in their peers, as they clearly understood the ideas the writers were aiming for, rather than a machine program like Grammarly, as stated by Students 1, 3, and 8. As a result, most of the feedback from their peers related to coherence, clarity, cohesion, and organization were more helpful for academic writing works. Even if there were misunderstandings, the writers could easily clarify and double-check with their friends, making the feedback more detailed, reliable, and constructive.

Furthermore, according to Student 2, 4, and 7, thanks to the method, they could find their weaknesses and areas for improvement in global and local fields while Grammarly was more likely to provide them with feedback in local fields. They also found the feedback to be a valuable learning resource in terms of grammar, ideas, content, organization, and vocabulary.

Lastly, the uniqueness of peer comments was a noteworthy factor. Student 8 highlighted that these comments were based on their peers' personal observations and understanding, leading to different viewpoints about her writing. This diversity of perspectives allowed writers to gain fresh insights into their own works.

Why did you choose Grammarly feedback?

One of the primary reasons that Grammarly outperformed peer feedback was immediate and automated suggestions. Students 6 and 9 highlighted the convenience of receiving instant corrections by simply pasting their text into Grammarly. This eliminated the waiting time of 3 to 4 days that peer feedback required, which was time-saving and suitable for those who prefer efficiency and a seamless writing experience.

The next factor was unbiased evaluation, as stated by Student 9. As Grammarly suggestions are based on predefined grammar rules, it was able to offer a standardized evaluation unlike peer feedback, which was influenced by personal biases or subjective opinions. Therefore, Grammarly could provide a more impartial assessment of the work.

Last but not least, in addition to grammatical corrections, Grammarly could offer suggestions on improving clarity, conciseness, and tone to improve the overall quality of the writing.

Discussion

Question 1: What are the perceptions of EFL postgraduate learners about feedback from peers to enhance their writing quality?

Regarding this question, Table 2 shows that 90% of the interviewees shared positive or very positive responses towards the method in many aspects, which has been investigated in the study of Ebadi and Rahimi (2017). Most of the EFL postgraduate students agreed that by implementing peer feedback into writing, the students got effective and constructive feedback rather than surface-level feedback from Grammarly; sometimes, their friends helped them

eliminate irrelevant ideas. Hence, the learners could improve their writing competence and the quality of their writing in terms of context organization and cohesion. The findings align with those of Huisman et al. (2018), Pham et al. (2020), and Latifi et al. (2023). By discussing with their classmates vague comments, the EFL postgraduate students could develop their critical thinking and foster a more analytical and reflective approach to their own writing, as in correspondence with the studies of Yang et al. (2006), Ekahitanond (2013), Novakovich (2016), and Vo (2022). In Vo's study (2022), the author explained that discussing those comments with friends motivates them to communicate effectively and helps students understand the problems clearly.

Furthermore, the students highlighted that peer feedback offered them a chance to expand their knowledge horizons. Through their peers' feedback, they were able to discover new vocabulary or writing styles and identify areas of weakness in their writing abilities. This observation aligned with the findings of Yang (2016), Kuyyogsuy (2019), and Bui et al. (2021), which also emphasized the positive influence of peer feedback on vocabulary acquisition and self-awareness of writing deficiencies. Consequently, peer feedback was regarded as a valuable and beneficial editing activity in writing classes.

On the other hand, the EFL postgraduate students viewed aspects of peer evaluation negatively. One significant concern was the quality of feedback received from unenthusiastic peers, who often provided vague, unconstructive, and unhelpful comments. This issue undermined the overall effectiveness of the feedback process, and this problem is aligned with the study of Vo (2022). Additionally, the EFL students expressed dissatisfaction with the limited time required to provide feedback. The time constraint restricted their ability to offer thoughtful and detailed comments, as supported by the research conducted by Rollinson (2005) and Kuyyogsuy (2019). Furthermore, the presence of varying levels of expertise and knowledge among peers could lead to differences in perspectives. Consequently, an idea that may seem suitable to one individual could be perceived as irrelevant by others, potentially resulting in conflicts and disagreements, as highlighted by Kuyyogsuy (2019). Finally, in cases where the students had close relationships with their peers or wished to maintain harmony within the class, they tended to provide biased comments to avoid making the writers, correlating with the research of Kunwongse (2013), Kuyyogsuy (2019), and Vo (2022).

Question 2: What are the perceptions of EFL postgraduate learners about feedback from Grammarly to enhance their writing quality?

In reference to the figures collected from the interview, the EFL postgraduate learners expressed favorable views regarding the integration of Grammarly in their academic writing class. Most learners praised Grammarly for its ability to provide instant and automated feedback, making the feedback process faster and more convenient. This result was also aligned with the result of previous papers by Wilson and Czik (2016), Fahmi and Cahyono (2021), and Dewi (2022), which found that Grammarly supported users to save their time in the revision stage; therefore, the writers themselves had more time to edit carefully before submitting their works. Meanwhile, the platform was able to provide detailed and helpful corrections, particularly on grammar and linguistics aspects, focusing mainly on local revisions, it also offered suggestions about clarity, conciseness, and tone improvement, which reduced the number of errors in their writing (Wilson & Czik, 2016; Ghufron & Rosyida, 2018; Fitriana & Nurazni, 2022; Dewi, 2022; Astuti et al., 2023). Besides that, the scoring system was a standout feature of the program that received high satisfaction from users. It allowed the EFL postgraduate students to access the quality of their work and edit it for better grades, which correlates with the findings of Astuti et al. (2023). The EFL postgraduate students also regarded the website well in terms of

accessibility and availability. This finding shares the same view with O'Neill and Russell (2019), Dewi (2022), Fitriana and Nurazni (2022), and Astuti et al. (2023), all of whom indicate that the app can be accessed anywhere and on any electronic device such as phones, computers, or laptops as well as its integration with Microsoft Word. The last feature, but also the most important one, was the unbiased evaluation, which significantly elevated the quality of the students' writing works, which corresponds to the statement of Astuti et al. (2023).

Nevertheless, misleading feedback contributed to reducing the level of trust of the users towards the program; those comments changed the intended meaning and intention completely, causing frustration among the EFL postgraduate students. The finding is in line with previous papers of Nova and Lukmana (2018), O'Neill and Russell (2019), and Fahmi and Cahyono (2021). Furthermore, when dealing with longer texts, Grammarly showed limitations in detecting errors, especially in citation, which subsequently affected the overall writing quality and this result is also highlighted in the studies of Nova and Lukmana (2018), O'Neill and Russell (2019), and Astuti et al. (2023). Grammarly is also proved to be less efficient in providing suggestions in terms of context improvement, restructuring ideas, and understanding some specialized terms, as also highlighted in the studies of Ghufron & Rosyida (2018), Ghufon (2019), Javier (2022), and Astuti et al. (2023). Lastly, it is noticeable that Grammarly predominantly gave feedback on local aspects rather than global ones, which the students at the high proficiency level did not highly appreciate.

Question 3: Between Grammarly and peer feedback, which method is more effective according to EFL postgraduate students?

In terms of the third question about the preferred method, the figures indicated that over two-thirds of the EFL postgraduate students gravitated towards peer feedback activities. The findings show that the EFL postgraduate students valued global revisions of peer feedback more than those local comments offered by Grammarly. The current finding corresponds with that of Huisman and co-authors (2018), who discovered that students enrolling in academic writing classes tend to prefer explanatory feedback more than analytical feedback. The top reason for this is that the students themselves could handle grammatical, lexicon, or spelling errors while Grammarly was found to be less efficient in adjusting organization and content (Ghufron & Rosyida, 2018). In other words, the higher English proficiency levels are the lower expectations for local revisions.

Conclusion

The primary objectives of the research are to investigate the perceptions of the EFL postgraduate learners on the integration of peer feedback and Grammarly feedback in improving their academic writing skills, as well as to determine which approach is more effective. Through employing the interview method, the study has successfully unveiled that the implementation of peer feedback and Grammarly feedback in the Research Writing course obtained good responses from the EFL postgraduate students. The findings demonstrate that the postgraduate students particularly valued peer feedback for its ability to facilitate global revision, foster critical thinking development, enhance their overall knowledge, etc. On the other hand, Grammarly feedback was commended for its real-time feedback, error reduction, easy accessibility and availability, etc.

It is also pointed out that for the EFL postgraduate students, peer feedback totally outperformed Grammarly feedback, primarily due to their strong appreciation for global revision as opposed to the focus on local revision. Another reason was that with peer feedback, the students could

discuss comments with their friends, while they could not do that with Grammarly.

Recommendations

We believe that English teachers should consider integrating peer feedback into their writing class so as to save time in accessing students' works and foster critical thinking skills among students. However, it is essential to consider students' English levels, especially when it comes to undergraduate students, due to the fact that different proficiency levels may need their own approach. For example, with students at A1 to B1 CEFR levels, instead of giving a whole text and asking them to give feedback to their friends, it is more commendable for teachers to break down the text into small sections so that the students can easily point out their friends' mistakes and then provided feedback. While with high-proficiency students (B2 to C2 CEFR levels), teachers can totally provide learners with a whole text and then ask them for review. Regardless of the students' proficiency levels, it is important to provide a checklist that outlines the assessment and evaluation criteria for various aspects such as task achievement, coherence and cohesion, vocabulary, and grammar. This checklist serves as a helpful tool for students to provide feedback to their peers in a more detailed and precise manner. Students can focus on specific criteria by referring to the checklist, ensuring that their feedback covers the necessary elements.

Besides that, with students at lower English proficiency levels - from A1 to B1 CEFR levels, English teachers can consider implementing Grammarly into their class due to the fact that the program can support them in terms of local revision so that the teachers can allocate more time to instruct students about how to organize ideas in a paragraph or essay and other global aspects. By tailoring the feedback approach to the student's language proficiency levels, educators can effectively enhance their students' writing skills and overall learning experience.

Nevertheless, it is also essential to consider the amount of time allocated for peer response activities. As the findings show, limited time can negatively impact the quality of peer feedback. Indeed, by taking into account the complexity of each writing task and their students' proficiency levels, English teachers can make informed decisions about setting appropriate time limits.

To future researchers, it might be beneficial for future studies to explore other contextual factors that could influence student's perceptions of peer feedback and Grammarly feedback. Factors such as cultural background, prior writing experiences, and individual learning preferences could potentially shape student preferences for specific feedback methods. Moreover, other researchers can conduct a study on the combination of both peer evaluation and Grammarly feedback on students' writing abilities. Investigating the impact of this combined approach on students' writing outcomes and perceptions would provide valuable insights for educators seeking effective feedback strategies. Additionally, investigating which method is more effective than the other in terms of developing students' writing competence and overall academic performance could be another good idea for future studies. Finally, it is also highly recommended for future researchers to diversify their study samples by including students with lower English proficiency levels.

Limitations

Despite the promising findings obtained in this study, it is crucial to acknowledge certain limitations in order to lay the groundwork for more rigorous and comprehensive research in the future. One primary limitation is the relatively small sample size in this investigation, which comprised only 10 participants from a Master's class. Therefore, it is imperative for future studies to expand the sample size. Another aspect worth considering is the reliance solely on

interviews as the data collection method in this research. While interviews are valuable for obtaining in-depth responses from participants, it might be advantageous for future research to employ other methods such as questionnaires, observations, and experimental and control groups.

Acknowledgments

References

Aidil, T. M. (2021). Exploring Students' Perception On The Effectiveness Of "Grammarly Software" In Academic Writing. UIN Ar-Raniry.

Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian journal of psychological medicine*, *43*(1), 86-88. DOI: https://doi.org/10.1177/0253717620977000.

Asper, G., Faria, C., Serra, P., & Galvão, C. (2024). Peer feedback and learning: a case study with 8th-grade Portuguese students. *Education 3-13*, 1-18. DOI: https://doi.org/10.1080/03004279.2024.2364610

Astuti, D., Darmahusni, D., Sumarni, S., & Suseno, M. (2023, June). The Use of Grammarly in the Academic Writing of Undergraduate Students: Advantages, Weaknesses, and Challenges (Systematic Review). *English Language and Literature International Conference (ELLiC) Proceedings*, *6* (1), 593-607.

Bridwell, L. S. (1980). Revising strategies in twelfth grade students' transactional writing. *Research in the Teaching of English, 14*, 197-222. DOI: https://doi.org/10.58680/rte198015795.

Bui, T. X. T., Ha, Y. N., Nguyen, T. B. U., Nguyen, V. U. T., & Ngo, T. C. T. (2021). A Study on Collaborative Online Learning among EFL Students in Van Lang University (VLU). *AsiaCALL Online Journal*, *12*(3), 13. Retrieved from https://asiacall.info/acoj/index.php/journal/article/view/32.

Burgess, A. W., Roberts, C., Black, K. I., & Mellis, C. (2013). Senior medical students perceived ability and experience in giving peer feedback in formative long case examinations. *BMC Medical Education, 13*(1), 1–5. DOI: https://doi.org/10.1186/1472-6920-13-79.

Butler, J. A., & Britt, M. A. (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication, 28*(1), 70-96. DOI: https://doi.org/10.1177/0741088310387891.

Chen, C. F. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology,12*(2), 94-112. DOI: https://doi.org/10125/44145.

Cotos, E. (2011). Potential of automated writing evaluation feedback. *Calico Journal*, *28*(2), 420-459. DOI: http://dx.doi.org/10.11139/cj.28.2.420-459.

Crane, T. (2005). What is the problem of perception?. *Synthesis Philosophica*, *20*(2), 237-264.

Crossley, S., & McNamara, D. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. *Proceedings of the Annual Meeting of the Cognitive Science Society, 33* (33), 101.

Dang, T. H. N. (2024). EFL Students' Perceptions of Peer Feedback in Writing Classes at a University in HCM City. *International Journal of Language Instruction*, *3*(2), 18–28. DOI: https://doi.org/10.54855/ijli.24322.

Dewi, U. (2022). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura*, *16*(2), 155-161. DOI: https://doi.org/10.21512/lc.v16i2.8315.

Dewi, U. (2023). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura*, *16*(2), 155-161. DOI: https://doi.org/10.21512/lc.v16i2.8315.

Ebadi, S., & Rahimi, M. (2017). Exploring the impact of online peer-editing using Google Docs on EFL learners' academic writing skills: A mixed methods study. *Computer Assisted Language Learning, 30*(8), 787-815. DOI: https://doi.org/10.1080/09588221.2017.1363056.

Efron, R. (1969). What is perception?. *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*. 137-173. DOI: https://doi.org/10.1007/978-94-010-3378-7_4.

Ekahitanond, V. (2013). Promoting university students' critical thinking skills through peer feedback activity in an online discussion forum. *Alberta Journal of Educational Research*, *59*(2), 247-265. DOI: https://doi.org/10.11575/ajer.v59i2.55617.

Epstein, W., West, L.J., & Dember, W.N., (2023). *Britannica: Psychology and Mental Health*. Retrieved from https://www.britannica.com/topic/perception.

Fahmi, M. A., & Cahyono, B. Y. (2021). EFL students' perception on the use of Grammarly and teacher feedback. *JEES (Journal of English Educators Society)*, *6*(1), 18-25. DOI: https://doi.org/10.21070/jees.v6i1.849.

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching*, *5*(1), 65-78. DOI: https://doi.org/10.31002/metathesis.v5i1.3519.

Fitriana, K., & Nurazni, L. (2022). Exploring English Department Students' Perceptions on Using Grammarly to Check the Grammar in Their Writing. *Journal of English Teaching*, *8*(1), 15-25. DOI: https://doi.org/10.33541/jet.v8i1.3044.

Fitriana, K., & Nurazni, L. (2022). Exploring English Department Students' Perceptions on Using Grammarly to Check the Grammar in Their Writing. *Journal of English Teaching*, *8*(1), 15-25. DOI: https://doi.org/10.33541/jet.v8i1.3044.

Freiberg, H. J. (Ed.). (1999). *School climate: Measuring, improving, and sustaining healthy learning environments*. Psychology Press.

Gaynor, J. W. (2020). Peer review in the classroom: Student perceptions, peer feedback quality and the role of assessment. *Assessment & Evaluation in Higher Education*, *45*(5), 758-775. DOI: https://doi.org/10.1080/02602938.2019.1697424.

Ghadirian, H., Fauzi Mohd Ayub, A., & Salehi, K. (2018). Students' perceptions of online discussions, participation and e-moderation behaviours in peer-moderated asynchronous online discussions. *Technology, Pedagogy and Education*, *27*(1), 85-100. DOI: https://doi.org/10.1080/1475939X.2017.1380695.

Ghufron, M. (2019, July). Exploring an automated feedback program 'Grammarly And teacher corrective feedback in EFL writing assessment: Modern vs. traditional assessment. *ELLiC 2019: Proceedings of the 3rd English Language and Literature International Conference, ELLiC, 27th April 2019, Semarang, Indonesia,* 307.

Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, *12*(4), 395-403. DOI: https://doi.org/10.21512/lc.v12i4.4582.

Grammarly. (2020). *About Grammarly*. Retrieved from https://support.grammarly.com/hc/enus/categories/115000018611-About-Grammrly

Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1625.

Hayes, J. R. (2000). A new framework for understanding cognition and affect in writing. *Perspectives on writing: Research, theory, and practice*, 6-44.

Ho, P. V. P., Thien, N. M., Ly, H. H., & Vy, N. N. H. (2020). The Practical Perceptions of Vietnamese Lecturers and Students Towards Written Peer Feedback. *International Journal of English Linguistics, 10*(6), 2006. DOI: https://doi.org/10.5539/ijel.v10n6p347.

Hockly, N. (2019). Automated writing evaluation. *ELT Journal, 73*(1), 82-88. DOI: https://doi.org/10.1093/elt/ccy044.

Huang, H. W., Li, Z., & Taylor, L. (2020, May). The Effectiveness of Using Grammarly to Improve Students' Writing Skills. *Proceedings of the 5th International Conference on Distance Education and Learning*, 122-127. DOI: https://doi.org/10.1145/3402569.3402594.

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, *43*(6), 955-968. DOI: https://doi.org/10.1080/02602938.2018.1424318.

Hutt, S., DePiro, A., Wang, J., Rhodes, S., Baker, R.S., Hieb, G., Sethuraman, S., Ocumpaugh, J., Mills, C. (2024). Feedback on Feedback: Comparing Classic Natural Language Processing and Generative AI to Evaluate Peer Feedback. *Proceedings of the 14th Learning Analytics and Knowledge Conference* (Kyoto Japan, 2024), 55–65. DOI: https://doi.org/10.1145/3636555.3636850.

Hyland, K. (2003). Second language writing. Cambridge: Cambridge University Press.

Iftanti, E. (2016). Improving students' writing skills through writing journal articles. *IAIN Tulungagung Research Collections*, *8*(1), 1-22. DOI: https://doi.org/10.21274/ls.2016.8.1.1-22.

Irvin, L. L. (2010). What is academic writing. *Writing spaces: Readings on writing*, *1*, 3-17.

Javier, D. R. (2022). App Review Using Tech Tools for Academic Writing:" Grammarly" as a Pedagogical Tool. *MEXTESOL Journal, 46*(2), 120.

Kasch, J., van Rosmalen, P., Löhr, A., Klemke, R., Antonaci, A., & Kalz, M. (2021). Students' perceptions of the peer-feedback experience in MOOCs. *Distance Education, 42*(1), 145-163. DOI: https://doi.org/10.1080/01587919.2020.1869522.

Kunwongse, S. (2013). Peer feedback, benefits and drawbacks. *Thammasat Review*, *16*(3), 277-288.

Kuyyogsuy, S. (2019). Promoting Peer Feedback in Developing Students' English Writing Ability in L2 Writing Class. *International Education Studies*, *12*(9), 76-90. DOI: https://doi.org/10.5539/ies.v12n9p76.

Lailika, H. I. (2019). *Students' Peceptions of The Use of Grammarly as an Online Grammar Checker in Thesis Writing*, *5* (1), 366-371. DOI: https://doi.org/10.22460/project.v5i2.p366-371.

Latifi, S., Noroozi, O., & Talaee, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, *31*(2), 655-669. DOI: https://doi.org/10.1080/10494820.2020.1799032.

Le, P. T. N. (2023). The Effectiveness of and Students' Perceptions of Peer Feedback: A Vietnam Action Research Project. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *15*(1), 12-14. DOI: https://doi.org/10.21659/rupkatha.v15n1.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer Nature.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the recent literature. *Personnel Psychology, 67*, 241–293. DOI: https://doi.org/10.1111/peps.12052.

Liu, J., & Hansen, G. J. (2005). *Peer response in second language writing classroom*. Ann Arbor, MI: University of Michigan Press.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 1*1(3), 279–290. DOI: https://doi.org/10.1080/13562510600680582.

McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Technical report, Institute for Intelligent Systems, University of Memphis, Memphis, TN.

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in human behavior*, *117*(1), 10-12. DOI: https://doi.org/10.1016/j.chb.2020.106658.

Noroozi, O., & Mulder, M. (2017). Design and evaluation of a digital module with guided peer feedback for student learning biotechnology and molecular life sciences, attitudinal

change, and satisfaction. *Biochemistry and Molecular Biology Education, 45*(1), 31–39. DOI: https://doi.org/10.1002/bmb.20981.

Nova, M., & Lukmana, I. (2018). The detected and undetected errors in automated writing evaluation program's result. *English Language and Literature International Conference, 7*(2), 120-126.

Novakovich, J. (2016). Fostering critical thinking and reflection through blog-mediated peer feedback. *Journal of Computer Assisted Learning*, *32*(1), 16-30. DOI: https://doi.org/10.1111/jcal.12114.

NURZAKIAH, A. (2021). *The Analysis of Students Perception of Online English Learning During Pandemic Covid 19 at The Eleventh Grade Students of SMAN 1 SAPE in Academic Year 2020/2021.* Undergraduate thesis, Universitas Muhammadiyah Mataram.

O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time university students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1), 42-56. Retrieved from https://journal.aall.org.au/index.php/jall/article/view/591.

Oshima, A., & Hogue, A. (2007). *Introduction to academic writing*. New York: Pearson Longman.

Oxford Learner's Dictionaries. (n.d.). Perception. Oxford University Press. Retrieved from https://www.oxfordlearnersdictionaries.com/definition/english/perception.

Parra G, L., & Calero S, X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, *12*(2), 209-226. DOI: https://doi.org/10.29333/iji.2019.12214a.

Peus, C., Braun, S., & Frey, D. (2013). Situation-based measurement of the full range of leadership model - development and validation of a situational judgment test. *The Leadership Quarterly, 24*, 777–795. DOI: https://doi.org/10.1016/j.leaqua.2013.07.006.

Pham, N. T. (2020). Factors influencing interaction in an online English course in Vietnam. *VNU Journal of Foreign Studies, 36*(3). 10. DOI: https://doi.org/10.25073/2525-2445/vnufs.4562.

Pham, V. P. H., & Usaha, S. (2016). Blog-based peer response for L2 writing revision. *Computer Assisted Language Learning*, *29*(4), 724-748. DOI: https://doi.org/10.1080/09588221.2015.1026355.

Pham, V. P. Ho., Phung, L. T., Oanh, T. T., & Giao, N. Q. (2020). Should peer e-comments replace traditional peer comments? *International Journal of Instruction, 13*(1), 295–314. DOI: https:// doi.org/10.29333/iji.2020.13120a.

Pratama, P., & Arriyani, N. (2021). Descriptive Text Writing: Peer Response Strategy and Students' Learning Motivation. *English Language in Focus (ELIF)*, *4*(1), 81-88. DOI: https://doi.org/10.24853/elif.4.1.81-88.

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies, 6*(9), 1884-1894. DOI: https://doi.org/10.17507/tpls.0609.23.

Ramage, J. D., Bean, J. C., & Johnson, J. (2003). *The Allyn & Bacon guide to writing*. Longman.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT journal*, *59*(1), 23-30. DOI: https://doi.org/10.1093/elt/cci003.

Ryan, T., & Henderson, M. (2017). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education, 43*(6), 880–892. DOI: https://doi.org/10.1080/02602938.2017.1416456.

Shermis, M., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Routledge.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. DOI: https://doi.org/10.1016/j.asw.2013.11.007.

Taghizadeh Kerman, N., Banihashem, S. K., Noroozi, O., & Biemans, H. J. (2022). The effects of students' perceived usefulness and trustworthiness of peer feedback on learning satisfaction in online learning environments. *8th International Conference on Higher Education Advances (HEAd'22), Universitat Politecnica de Valencia*. 265-268. DOI: http://dx.doi.org/10.4995/HEAd22.2022.14445.

Tai, J. H. M., Canny, B. J., Haines, T. P., & Molloy, E. K. (2015). The role of peer-assisted learning in building evaluative judgement: opportunities in clinical medical education. *Advances in Health Sciences Education 2015, 2*1(3), 659–676. DOI: https://doi.org/10.1007/S10459-015-9659-0.

Thi, N. K., Nikolov, M., & Simon, K. (2022). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 1-16. DOI: https://doi.org/10.1080/17501229.2022.2122476.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Vo, T. T. M. (2022). EFL Students' Attitudes Towards Teacher Correction and Peer Correction in Writing Skills. *International Journal of Language Instruction*, *1*(1), 155–173. DOI: https://doi.org/10.54855/ijli.221113.

Wakabayashi, R. (2013). The Effects of the Peer Feedback Process on Reviewers' Own Writing. *English Language Teaching*, *6*(9), 177-192. DOI: http://dx.doi.org/10.5539/elt.v6n9p177.

Warschauer, M., & Ware P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language teaching research*, *10*(2), 157-180. DOI: https://doi.org/10.1191/1362168806lr190oa.

Wilson, J., & Andrada, G. N. (2016). Using automated feedback to improve writing quality: Opportunities and challenges. *Handbook of research on technology tools for real-world skill development*, 679-704. DOI: http://dx.doi.org/10.4018/978-1-4666-9441-5.ch026.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in Chinese EFL writing class. *Journal of Second Language Learning, 15*(3), 179-200. DOI: https://doi.org/10.1016/j.jslw.2006.09.004.

Yang, Y. F. (2016). Transforming and constructing academic knowledge through online peer feedback in summary writing. *Computer Assisted Language Learning*, *29*(4), 683-702. DOI: https://doi.org/10.1080/09588221.2015.1016440.

Zeevy-Solovey, O. (2024). Comparing peers, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, *6*(3), 1-23. DOI: http://dx.doi.org/10.29140/tltl.v6n3.1482

Biodata

Yen Nhi Ha is an accomplished graduate student who holds an excellent degree from the Faculty of Foreign Languages at Van Lang University. Currently, she is working as a lecturer at the Language Institute of Van Lang University. About her research background, Nhi has demonstrated her scholarly abilities by presenting her research papers at different conferences. Her research interests primarily revolve around the integration of technology in language teaching and learning, online education, and the development of English language proficiency.

Ngoc Phuong Ho graduated with a BA in English Language in 2022. She has worked as an ESL teacher for 2 years. Currently, she is pursuing an MA in English Language at the Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam. Her research interests encompass Quality assurance in Vietnamese higher education, Translation, and Pragmatics.

The Practice of ChatGPT in English Teaching and Learning in Vietnam: A Systematic Review

Pham Manh Tri[1,2], Cao Thi Xuan Tu[2*]

[1] Institute of Postgraduate Education, Van Lang University, Ho Chi Minh City, Vietnam

[2] Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam

[*] Corresponding author's email: tu.ctx@vlu.edu.vn

[*] https://orcid.org/0000-0001-7871-6429

https://doi.org/10.54855/ijte.25513

ABSTRACT

The purpose of the research is to explore the practice of using ChatGPT in teaching and learning English in the context of Vietnam based on a systematic review of 12 selected studies from 2023 to 2024, which were retrieved from journals such as International Journal of TESOL & Education, AsiaCALL Online Journal, Teaching English with Technology, European Journal of Alternative Education Studies, International Journal of Learning, Teaching and Educational Research, International Journal of Language Instruction, and Kognisi: Jurnal Ilmu Keguruan via Google Scholar. The detailed analysis focuses on the benefits, challenges, and issues of the use of ChatGPT in English teaching and learning in the context of Vietnam. The findings indicate that the deployment of ChatGPT greatly contributes to language education, namely the facilitation of students' and teachers' work. Nevertheless, there remain unsolved issues regarding academic dishonesty, plagiarism, sole reliance upon the tool's function, the appropriateness of the information it offers. It is suggested that there should be training or guidance for teachers and students to use ChatGPT properly, balanced integration of the tool with teachers' consideration, and so forth. The review article provides valuable references and pedagogical recommendations for future research papers about ChatGPT's use in English language education.

Keywords: systematic review, ChatGPT, English education, artificial intelligence (AI)

Introduction

In this day and age, the invention of technology is considered a great impact on education (Pham et al., 2022). Besides, the development of artificial intelligence (AI) contributes to educational efficiency and effectiveness, personalized and global learning, and intelligent content creation (Montenegro-Rueda et al., 2023). Nowadays, technology or AI plays a vital role in learning foreign languages such as English (Pham & Le, 2024; Pham et al., 2023) and creating classroom

activities (Tri et al., 2023). When it comes to AI, ChatGPT has become a viral AI chatbot after its release because it supports users' work with its human-like responses (Lo, 2023). Since ChatGPT has the ability to deal with diverse topics and has a strong vocabulary base, it is regarded as a valuable tool for language teaching and learning (Cotton et al., 2023).

Notwithstanding ChatGPT's success, it still has challenges to English language education. The issue regarding academic integrity such as plagiarism is unsolved (Iskender, 2023; Cotton et al., 2023). In addition, instructors are afraid that learners will never use their own words to complete their tasks (Cotton et al., 2023). There is no assurance the AI tool also provides wrong knowledge (Nguyen, 2024). Overuse of AI results in the decline of critical thinking skills and cheating, so educators must be conscious of concerns to make sure that it is used safely and properly in educational contexts (Kostka & Toncelli, 2023).

In the context of Vietnamese education, ChatGPT has become a new trend recently. The tool facilitates students' tasks (Nguyen & Tran, 2023; Ho, 2024), provides opportunities to practice using English (Hoang et al., 2023; Thao et al., 2023; Nguyen, 2023; Nguyen et al., 2024; Nguyen & Tran, 2023), and improves learner autonomy (Thao et al., 2023). ChatGPT also alleviates teachers' workload pressure in terms of evaluation (Nguyen, 2024; Nguyen & Tran, 2023; Nguyen, 2023). Using ChatGPT for English learning and teaching is quite new, leading to unforeseen and unsolved problems, namely academic dishonesty (Ho, 2024; Cong-Lem et al., 2024), over-dependence on the tool, quality, accuracy, appropriateness, privacy, ethics (Thao et al., 2023), training (Nguyen, 2023; Nguyen, 2024; Tran & Tran, 2024), etc.

Using ChatGPT for English education in the context of Vietnam is a topic attracting researchers' and educators' attention. Numerous researchers and educators highlighted the role of ChatGPT in English language education. Nonetheless, there are no systematic reviews discussing utilizing ChatGPT for teaching and learning in Vietnamese English as a Foreign Language (EFL) classrooms. Conducting a systematic review discussion about ChatGPT, especially in the context of Vietnam, can help deeply understand aspects in terms of linguistics, culture as well as how this tool works in a specific setting. Besides, it is crucial that learners and instructors have a broad knowledge of the application of ChatGPT to ensure its proper use and maintain the quality of education (Lozano & Fontao, 2023). The findings of the literature on the use of ChatGPT in EFL teaching and learning in Vietnam have to be reviewed to summarize what researchers and educators need for future studies and practices in this context. For this reason, the main purpose of the review article is to examine the existing literature on the application of ChatGPT in EFL teaching and learning in Vietnam with a view to exploring its advantages and disadvantages as well as recommendations from the previous studies.

Literature Review

Systematic review

The purpose of a systematic review is to unite data to address a predetermined question (Pollock & Berge, 2018). This involves identifying all primary research pertinent to the review question, critically evaluating the research, and synthesizing the results (Gough et al., 2017). Systematic reviews integrate data from various papers to generate a new, combined finding or conclusion,

or perhaps they compile dissimilar sorts of evidence to investigate or account for meanings (Snilstveit, 2012). Traditional reviews lack a reproducible or formal method for appraising the impact of a treatment, counting its size and accuracy, so a more structured approach is essential (Egger et al., 2001; Tricco et al., 2011). The systematic review, which is also called "research synthesis", strives to offer a comprehensive, unbiased synthesis of numerous related research papers in one document (Egger et al., 2001; Khan et al., 2003; Tricco et al., 2011). While it shares multiple characteristics with a literature review, such as summarizing knowledge from a body of literature, a systematic review aims to uncover all evidence associated with a question, focusing on research reporting data in place of concepts or theories (Averis & Pearson, 2003).

ChatGPT

ChatGPT, launched in November 2022, is described as a "state-of-the-art chatbot" according to Hong (2023). Zhai (2022) defines this tool as a chatbot offering human-like and open-ended conversations for specific purposes to discuss a variety of subjects. It can compare current data in order to give the most appropriate answers to different types of queries (Nguyen, 2023).

The role of ChatGPT in English language education

After ChatGPT's release, it attracted the attention of researchers, educators and learners across the globe because of its positive impact on education. Recently, using ChatGPT in language education has become a viral topic (Tran & Tran, 2023). The utilization of this AI enables language students to participate in real-life conversations improving their fluency via individualized exercises and promote autonomous learning by permitting practicing their target language skills on their own (Tran & Tran, 2023). ChatGPT's instant comments and authentic conversations can advance students' speaking and writing skills (Aljanabi, 2023). In addition, students are given opportunities to enhance their language skills with a learner-centered approach through meaningful interactions (Yang & Kyun, 2022).

Advantages of ChatGPT in English language education

ChatGPT is described as a useful tool for improving language teaching methods, giving comments at once and answering a wide range of questions (Mohamed, 2023). The AI tool helps to get access to innumerable authentic linguistic resources (Hong, 2023; Mai et al., 2024). Kohnke et al. (2023) reported that ChatGPT can account for difficult concepts and provide word usage, examples, and error correction through detailed explanations, translations, and so on with the aim of language enhancement. Thanks to the AI tool, students are more likely to increase their motivation in their English language learning journey (Ali et al., 2023). Additionally, Hong (2023) and Kohnke et al. (2023) agreed that ChatGPT makes learning experiences more interesting and personalized.

A large number of published papers also discussed the benefits and opportunities for the use of ChatGPT in English education in Vietnam. For students, ChatGPT boosts their engagement and interest in language tasks (Nguyen & Tran, 2023). Additionally, challenges of learning English can be tackled because ChatGPT assists students with vocabulary acquisition, translation, grammar checking, paraphrasing, and so forth (Ho, 2024). Students have chances to practice English to communicate (Hoang et al., 2023) and write academically (Thao et al., 2023; Nguyen, 2023; Nguyen et al., 2024; Nguyen & Tran, 2023) by means of ChatGPT, leading to learners'

promoted autonomous learning (Thao et al., 2023). Furthermore, the tool supports teachers' work in assessments (Nguyen, 2024; Nguyen & Tran, 2023) and test design (Nguyen, 2023). Nguyen and Tran (2023) also confirmed a clear resemblance between the grades assigned by ChatGPT and the teacher. Therefore, there should be collaboration between instructors and AI to improve teaching methods and evaluation (Tran & Tran, 2024; Thao et al., 2023). It is noted that ChatGPT is not able to replace the role of instructors, but they can employ it to support their teaching designs with their double-check (Mai et al., 2024).

Disadvantages of ChatGPT in English language education

Regardless of ChatGPT's benefits, this AI tool also causes many concerns for teaching and learning (Lo, 2023). It still leads to drawbacks affecting learners such as the decline of critical thinking skills (Mohamed, 2023). Stojanov (2023) revealed that students tend to have trouble using the tool effectively owing to their lack of essential skills and knowledge. In addition, ChatGPT is not a perfect tool containing misinformation and needs human considerations (Ali et al., 2023). In terms of academic integrity, students tend to use the AI tool to cheat in their learning process. For example, their own work contains AI-generated text without any originality of ideas (Sullivan et al., 2023).

Several papers also show a number of disadvantages and challenges related to the use of ChatGPT in the Vietnamese context. In terms of plagiarism issues, it is impossible to prevent students from the abuse of ChatGPT (Ho, 2024). There exist other concerns such as academic dishonesty (Cong-Lem et al., 2024), heavy dependence on AI, worries about the exactness and suitability of the content AI created, the limited ability to create something on students' own, privacy and ethical issues (Thao et al., 2023). Besides, There is a lack of training and instructions for using ChatGPT suitably (Nguyen, 2024; Tran & Tran, 2024).

Research gaps

The previous studies reported their successful results contributing to the field, but there remain limitations in the existing literature.

Issues concerning academic dishonesty still exist. Ho (2024) reported that guiding students to utilize ChatGPT is inadequate to cope with cheating issues. This researcher admitted that her research failed to explore an effective tool to detect plagiarism due to the abuse of the AI tool. This gap emphasizes the importance of future exploration of methods and useful AI content detectors to mitigate cheating and plagiarism when AI tools such as ChatGPT are used in educational settings.

Because ChatGPT is new to educators, they may struggle to get accustomed to the application of this tool to teaching. According to Nguyen (2023), some instructors still need time to familiarize themselves with ChatGPT though they got hands-on experience with the integration of the tool and their teaching practice. This author claimed that the shortage of training limits teachers' ability to employ this tool effectively. Nguyen (2024) also agreed that lack of training for instructors to exploit ChatGPT is a current concern. Drawing from this gap, one can deduce that there is a necessity for teachers to participate in professional programs about ChatGPT use for language teaching and learning.

Because of the small sample size, it is difficult for several researchers to generalize their conclusions. Nguyen et al. (2024) and Thao et al. (2023) found that their study is unlikely to represent the student population at universities in all regions of Vietnam. Similarly, it is impossible to fully understand the teacher population because of the small number of participants (Yen et al., 2024). Hieu and Thao (2024) had difficulty in matching diverse cultural or educational contexts because of their paper's specific research setting. Moreover, this limitation prevents Nguyen and Tran (2023) from heightening awareness of teachers' role in teaching and assessment by means of the tool when learners can utilize it as teachers do. The limited number of teacher participants also affects teachers' perceptions and experiences in general (Nguyen, 2023; Nguyen, 2024). These limitations underscore the need for further research to be conducted with larger samples to grasp deeper insights of both learners and teachers.

Numerous papers are solely dependent on qualitative approaches (e.g., Hieu & Thao, 2024; Nguyen, 2024; Nguyen & Tran, 2023; Thao et al., 2023; Tran & Tran, 2023; Yen et al., 2024) or quantitative approaches (e.g., Cong-Lem et al., 2024; Nguyen, 2024) instead of using mixed methods, impacting deep insights of the use of ChatGPT. It is underscored that using mixed-methods approaches can provide a better understanding of the utilization of this tool in educational settings.

There are no systematic reviews on the use of ChatGPT for English teaching and learning in the context of Vietnam; therefore, it is important to review and synthesize the findings of the previous studies to provide insights into the current state of research in this context. The researchers conducted this literature review paper to address the gaps, namely insufficient research on academic integrity matters, necessity for training, limited generalizability, and single-method approaches. This paper contributes to the development of knowledge and practice in this field and provides a broad understanding for educators, researchers, and policymakers to effectively incorporate ChatGPT within educational experiences. Specifically, this paper analyzes the literature on ChatGPT use for English education in Vietnam to explore its benefits, challenges, future trends, and emerging research areas. The discovery of the review helps answer the following research questions:

1. *What are the benefits and challenges of integrating ChatGPT with English education in Vietnam?*
2. *What future trends and emerging research areas can be identified in the application of ChatGPT for English language teaching and learning?*

Methods

Design of the Study

In order to thoroughly examine all relevant research, the authors conduct a systematic literature review using a clear and organized search plan that outlines what study will be included and excluded. This plan follows the guidelines set forth in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (Page et al., 2021). The researchers searched for relevant studies in education in many journals, namely *International*

Journal of TESOL & Education, *AsiaCALL Online Journal*, *Teaching English with Technology*, *European Journal of Alternative Education Studies*, *International Journal of Learning, Teaching and Educational Research*, *International Journal of Language Instruction*, and *Kognisi: Jurnal Ilmu Keguruan* mainly across Google Scholar. The researchers used a specific search strategy (detailed in Table 1) that looked for keywords in the titles, abstracts, and/or keywords of articles.

Search strategy

Table 1.

Search strategy

Topic	Search terms
ChatGPT Chatbot Artificial Intelligence	Language teaching and learning in Vietnam Impacts on language learners Practice of English teaching and learning

This search, conducted between April and May 2024, identified 20 initial records. Clear guidelines were defined to select studies directly related to the review's goals. The authors narrowed the search to articles published since 2022, which coincides with the launch of the specific AI chatbot developed by OpenAI.

Inclusion and exclusion criteria

Table 2.

Inclusion and exclusion criteria

|	**Inclusion Criteria**	**Exclusion Criteria**
Publication period	Published between 2022-2024	Published before 2022
Type of study	Theoretical and empirical research	Other research (review)
Language	English	Vietnamese or other languages
Context	Vietnam	Other regions
Research topic	English language teaching and learning	Education in general

Selection of studies

The Johanna Briggs Checklist (JBI) (Aromataris & Munn, 2020), which involves a set of criteria used to identify the rigor and validity of a study, is used to evaluate the methodological quality of selected research studies. The 20 selected studies were assessed using the JBI checklist in terms of aspects such as study design, participant selection, and data analysis. The following checklist was implemented to avoid any bias, and the selected studies had to meet at least four of the criteria:

· Is the research question clearly stated?

· Does the research explore how ChatGPT is used in language teaching and its effects?

· Are the methods used to collect data suitable for the research?

· Do the findings contribute valuable insights to the field of research?

· Do the authors' interpretations of the data justify their conclusions?

· Does the research suggest areas for further investigation?

After removing the studies that did not meet at least four criteria (n = 8), a total of 12 studies were eligible for review. Figure 1 demonstrates the flow chart of choosing studies according to the instructions of PRISMA (Page et al., 2021).

Figure 1.
The flow chart of the study selection process
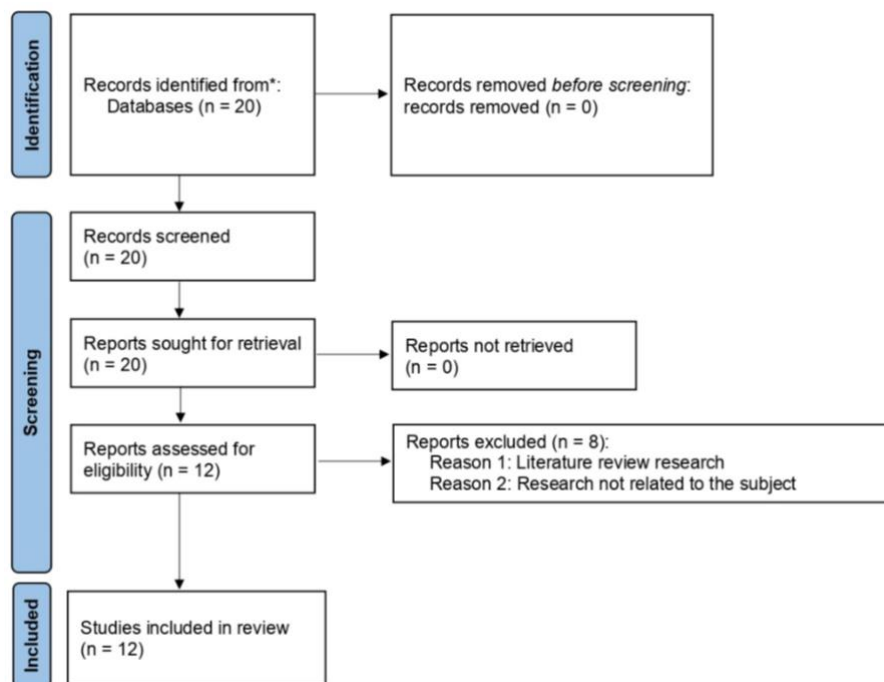
Data Extraction and Analysis

To answer the research questions, the researchers conducted an analysis that integrated quantitative and qualitative methods into the 12 papers. The quantitative analysis assisted in visualizing and comprehending common aspects concerning the subject via descriptive graphs. When it comes to the qualitative analysis, the VOSviewer 1.6.20. was employed to find out the main trends and impact of the research areas (Nandiyanto & Al Husaeni, 2021). A cluster has a wide variety of elements and sizes. The frequency of occurrence affects the size. The use of keywords in the articles is more popular, making sure that the circle is bigger (Mulyawati & Ramadhan, 2021). There is a list of criteria used for data analysis regarding the benefits of the challenges of using ChatGPT for English education, and the selected studies had to meet at least four of the criteria:

· Does the research have data on the benefits of the use of ChatGPT for English teaching?

· Does the research have data on the benefits of the use of ChatGPT for English learning?

· Does the research have data on the challenges of the use of ChatGPT for English teaching?

· Does the research have data on the challenges of the use of ChatGPT for English learning?

· Are the data of the research precise and clearly reported?

· Are the data of the research relevant to the context and significant for the research purpose?

· Are the data of the research valid and reliable?

Figure 2 describes the use of methods in the selected papers included in the review. A qualitative approach was utilized by most articles (50%) (n = 6). A quantitative approach was applied by two papers (17%) (n = 2). The other studies (33%) (n = 4) employed a mix of quantitative and qualitative methods.

Figure 2.
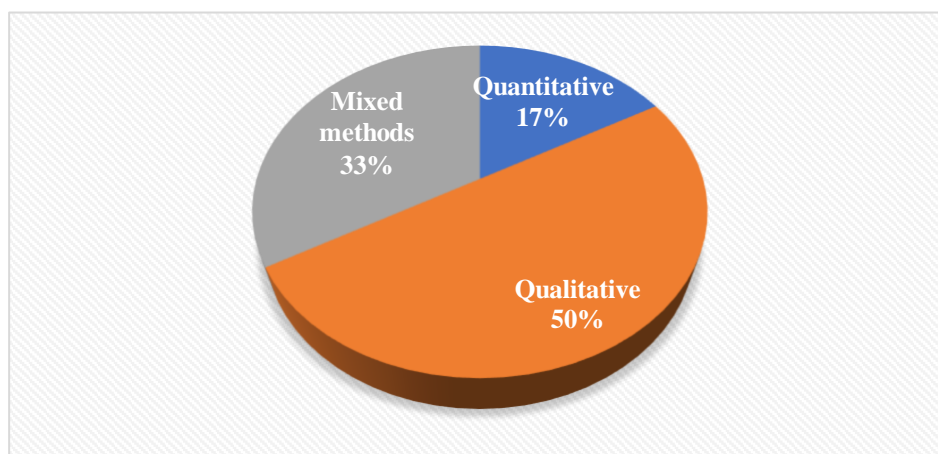
The methods of the selected studies (n = 12)

Table 3.

Research settings and methods of the 12 studies

No.	Author(s)	Study	Setting(s)	Journal	Methodology	Participant(s)
1	Ho (2024)	Using ChatGPT in English Language Learning: A Study on I.T. Students' Attitudes, Habits, and Perceptions	The University of Da Nang	International Journal of TESOL & Education	Quantitative	120 students
					Qualitative	10 students
2	Nguyen (2024)	University Teachers' Perceptions of Using ChatGPT in Language Teaching and Assessment	Universities	AsiaCALL Online Journal	Quantitative	43 teachers
3	Cong-Lem, Tran, & Nguyen (2024)	Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers	Universities	Teaching English with Technology	Quantitative	31 teachers
4	Nguyen, Ngoc, & Dan (2024)	EFL Students' Perceptions and Practices of Using ChatGPT for Developing English Argumentative Essay Writing Skills	Can Tho University	European Journal of Alternative Education Studies	Quantitative	100 students
					Qualitative	10 students
5	Yen, Thu, Thi, Tra,	University Teachers' Perceptions on the	Universities	AsiaCALL Online	Qualitative	15 teachers

	& Thuy (2024)	Integration of ChatGPT in Language Education Assessment: Challenges, Benefits, and Ethical Considerations		Journal		
6	Nguyen (2024)	Unraveling the Potential of ChatGPT: Investigating the Efficacy of Reading Text Adaptation	A university in Hanoi	AsiaCALL Online Journal	Qualitative	10 teachers
7	Hieu & Thao (2024)	Exploring the Impact of AI in Language Education: Vietnamese EFL Teachers' Views on Using ChatGPT for Fairy Tale Retelling Tasks	Universities	International Journal of Learning, Teaching and Educational Research	Qualitative	9 teachers
8	Nguyen & Tran (2023)	Exploring the Efficacy of ChatGPT in Language Teaching	The University of Da Nang	AsiaCALL Online Journal	Qualitative	1 teacher
9	Nguyen (2023)	The Application of ChatGPT in Language Test Design – The What and How	Universities and schools	AsiaCALL Online Journal	Quantitative	70 teachers
					Qualitative	5 teachers
10	Nguyen (2023)	EFL Teachers' Perspectives toward the Use of ChatGPT in Writing Classes: A Case Study at Van Lang University	Van Lang University	International Journal of Language Instruction	Quantitative	20 teachers
					Qualitative	10 teachers
11	Tran & Tran (2023)	Exploring the Role of ChatGPT in Developing Critical Digital Literacies in Language Learning: A Qualitative Study	Schools	AsiaCALL Online Journal	Qualitative	8 students & 3 teachers
12	Thao, Hieu, & Thuy (2023)	Exploring the Impacts of ChatGPT in EFL Writing: Student Perceptions of Opportunities and Challenges in Vietnamese Higher Education	Universities	Kognisi: Jurnal Ilmu Keguruan	Qualitative	20 students

Table 3 shows detailed information on the 12 selected studies, namely research settings and methods. Most of them were carried out in a variety of universities; however, Study 11 merely focused on schools, and Study 9 was conducted in both universities and schools. The results of the research papers are quantitative, qualitative, or a mix of them. There exist eight studies using only one method. Specifically, the findings of Study 2 and Study 3 are quantitative, and the ones of Study 5, Study 6, Study 7, Study 8, Study 11, and Study 12 are qualitative. Besides, Study 1, Study 4, Study 9, and Study 10 employed a mix of both quantitative and qualitative methods, making their data more valid (Spratt et al., 2004). With regard to the quantitative approach, Study 1 had the highest number of student participants (120 students), and Study 9 had the largest number of teacher participants (70 teachers) compared to other studies. With

respect to the qualitative approach, Study 12 had 20 participants considered the biggest number.

Findings and Discussion

As for the data found after reading the selected papers, Table 4 displays key findings related to the benefits and challenges of the use of ChatGPT in English education in Vietnam.

Table 4.

The benefits and challenges of the ChatGPT use for English teaching and learning

Benefits	Challenges
1. Personalized learning (Study 1, 2, 3, 4, 5, 6, 7, 8, 9, 11)	1. Quality and accuracy of AI-generated content (Study 1, 2, 3, 4, 6, 7, 8, 9, 11, 12)
2. Immediate assistance and feedback (Study 2, 4, 5, 6, 8, 10, 11)	2. AI over-reliance (Study 1, 3, 4, 5, 7, 10, 11, 12)
3. Promoted learner autonomy (Study 2, 5, 8, 11, 12)	3. Challenges in managing academic integrity (Study 1, 3, 4, 6, 8, 9)
4. Enhanced engagement (Study 3, 6, 7, 9, 12)	4. Ethical and privacy issues (Study 2, 3, 5, 6, 8, 12)
5. Teacher workload reduction (Study 1, 6, 8, 10)	5. Reduced human interaction (Study 1, 2, 6, 8, 11)
6. Automated assessment (Study 1, 3, 4, 12)	6. Lack of training (Study 2, 3, 5, 10)
7. Development of language skills (Study 1, 3, 4, 12)	7. Technical and resource limitations (Study 4, 7, 10, 11)
8. Development of critical thinking skills (Study 4, 7, 8)	8. Balancing ChatGPT with traditional teaching methods (Study 8, 11)
9. Development of creativity (Study 7, 9)	9. Decrease in students' creativity (Study 9, 12)
10. Access to authentic resources (Study 3, 11)	10. Cultural alignment challenges (Study 7)
11. Diversified teaching and learning strategies (Study 5, 11)	11. Negative impact on students' critical thinking skills (Study 9)
12. Reduced language anxiety (Study 3, 4)	12. Difficulty in responding to questions requiring higher cognitive thinking (Study 9)
13. Enhanced individualized language practice (Study 11)	13. Risk of promoting laziness (Study 8)
14. Enhanced learning outcomes (Study 10)	14. Challenges in monitoring and assessing ChatGPT use (Study 11)
15. Development of digital literacy skills (Study 8)	15. Potential to replace teachers (Study 1)
16. Development of soft skills (Study 2)	
17. Authentic language use (Study 9)	
18. Exposure to diverse language use (Study 12)	
19. Facilitated learning tasks (Study 1)	
20. Answers generated in formal examinations (Study 9)	
21. Language translation (Study 4)	
22. Collaborative future (Study 5)	
23. Exploration of unused features (Study 2)	

Based on Table 4, it is undeniable that ChatGPT plays a pivotal role in EFL classrooms in Vietnam. For students, their learning can be personalized thanks to this tool's immediate responses and comments, supporting their learner autonomy, facilitating their learning tasks, and increasing their interest. Besides, ChatGPT assists in boosting students' language skills like writing, critical thinking, creativity, digital literacy as well as soft skills. Regarding language skills, students can overcome language anxiety because they get access to authentic resources, individualized language practice, authentic language use, and language translation provided by ChatGPT. Moreover, learners benefit from exposure to a wide range of language uses and styles through their communication with this tool. The exposure can widen their horizons of language nuances and help them gain more a varied vocabulary. For teachers, ChatGPT can alleviate their workload, allowing them to pay more attention to deeper language comprehension among students. For instance, they can have their students' work and skills assessed automatically. The new learning and teaching experience can contribute to students' improved learning outcomes and diversify teaching and learning methods. Furthermore, many of ChatGPT's undiscovered features and abilities are supposed to be advantageous if applied and supported through appropriate training. Hence, there is a propitious future collaboration between AI tools like ChatGPT and traditional pedagogical approaches, which can revolutionize educational outcomes.

In spite of ChatGPT's benefits for English education in Vietnam, there are multiple drawbacks. Owing to the power of ChatGPT's functions, learners are more likely to over-rely on it. Thus, managing academic integrity is not an easy job, causing plagiarism issues and promoting laziness among students. Additionally, overusing this tool may reduce students' creativity, critical thinking, and human interaction, raising the possibility of replacing human instructors. It is tough to guarantee the quality and precision of ChatGPT's content because it is a master of all fields. The limitation will surely have effects on the quality of learning and this tool's dependability in various educational settings. For example, ChatGPT still has trouble answering complex questions requiring higher cognitive thinking. What this tool shows might not match the cultural contexts of language education. There are ethical issues such as data consent and privacy influencing trust in the tool. Instructors lack professional training to effectively combine ChatGPT with their teaching practices as well as deal with difficulties related to its use and technical and resource limitations, making it difficult for them to balance this tool with traditional teaching methods, and monitor and assess its use.

The research on ChatGPT's impact on language learning paints a complex picture. While it offers clear benefits for learners, such as improved vocabulary and writing skills, there are also concerns. Students may become overly reliant on AI, leading to a decline in critical thinking and plagiarism issues. For teachers, ChatGPT has the potential to reduce workload, but proper training is necessary to ensure effective use and address ethical concerns. The overall takeaway is that ChatGPT can be a valuable tool, but it needs to be integrated thoughtfully alongside traditional teaching methods to maximize its benefits for language learners while mitigating potential drawbacks. Further research is needed to explore the long-term impact of ChatGPT and develop solutions to plagiarism issues.

Figure 3.

Labeled bibliometric map
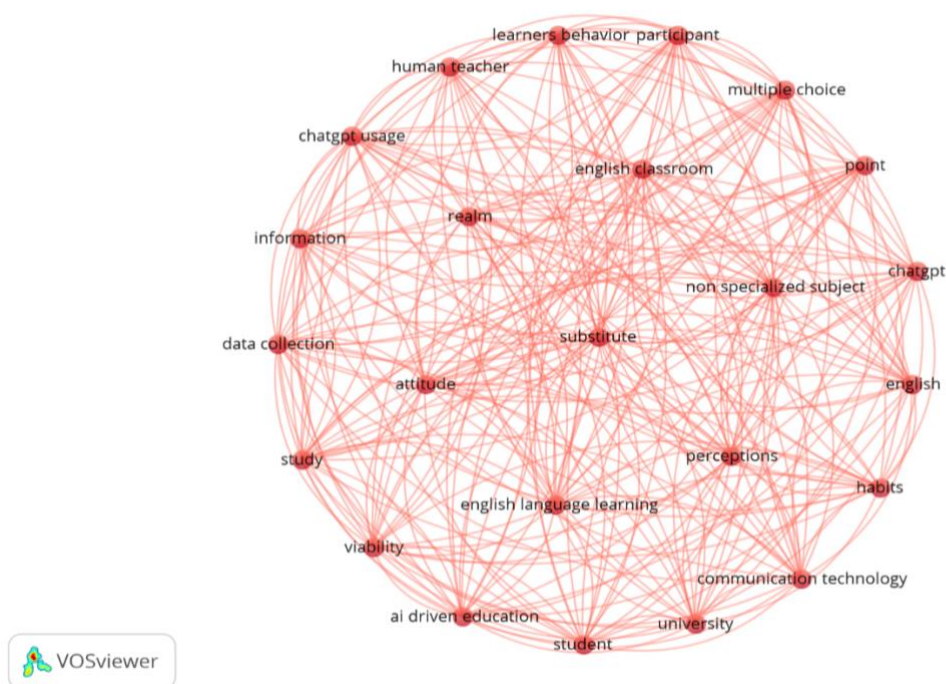
In order to discover the lines of research in the field, the researchers used VOSviewer's mapping. VOSviewer displays a cluster of 24 items (Figure 3): AI-driven education, attitude, ChatGPT, ChatGPT usage, communication technology, data collection, English, English classroom, English language learning, habits, human teacher, information, learners behavior, multiple choice, non-specialized subject, participant, perceptions, point, realm, student, study, substitute, university, and viability. This cluster is connected with research on human teachers' role in the AI era. ChatGPT is a type of technology that can be used to facilitate English language learning and teaching, but instructors are still responsible for going over its usage. It is essential that teachers adapt to innovation in their pedagogical practices.

Discussion

Research question 1: What are the benefits and challenges of integrating ChatGPT with English education in Vietnam?

The benefits of integrating ChatGPT with English education in Vietnam

1. Personalized Learning Environments

Using ChatGPT in Vietnamese EFL classrooms creates personalized learning environments suitable for students' needs, enhancing their learning experiences (Cong-Lem et al., 2024; Hieu & Thao, 2024; Ho, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Yen et al., 2024). For instance, learners can receive immediate learning support and assistance answering their questions on complex topics well, which is good for proofreading, brainstorming, and research

(Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Tran & Tran, 2023; Yen et al., 2024). Additionally, ChatGPT can customize enjoyable English language practice by employing conversational interactions (Cong-Lem et al., 2024; Tran & Tran, 2023), thereby boosting students' motivation and engagement in learning activities (Hieu & Thao, 2024; Ho, 2024; Nguyen, 2023; Nguyen, 2024). Notably, Thao et al. (2023) support the evidence by revealing that combining ChatGPT with EFL writing tasks increased engagement and interest, improving learners' participation and motivation.

2. Teacher Support and Efficiency

ChatGPT can save teachers' time, effort, and workload as it can supply personalized feedback for students and grade their work effectively, enabling teachers to focus on nurturing language comprehension among students (Ho, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Nguyen et al., 2024) and improve students' learning outcomes and satisfaction (Nguyen, 2023; Nguyen, 2024). In students' learning process, learning tasks can be simplified with the aid of ChatGPT's translation, grammatical error correction, summarization, and explanation with examples, leading to language development (Cong-Lem et al., 2024; Ho, 2024; Nguyen et al., 2024). However, there is a lack of human interaction and instructors' personalized feedback (Ho, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen, 2024; Tran & Tran, 2023).

3. Skill Development

When it comes to the development of skills and abilities, students can enhance reading skills (Ho, 2024), writing skills (Nguyen et al., 2024; Thao et al., 2023), speaking skills (Cong-Lem et al., 2023), learner autonomy (Ho, 2024; Thao et al., 2023; Tran & Tran, 2023; Yen et al., 2024), creativity and critical thinking skills (Hieu & Thao, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen, 2024; Nguyen & Tran, 2023). Concerning reading skills, ChatGPT generates authentic materials that contribute to learners' improved reading comprehension (Ho, 2024). Nguyen (2024) confirmed that ChatGPT can help learners adapt to reading materials based on their various reading abilities. Regarding writing skills, ChatGPT aids students in improving multiple aspects of their writing, namely vocabulary, grammar, organization, and style owing to supplying instant personalized feedback and recommendations assisting in revising their essays and academic papers (Ho, 2024; Nguyen, 2023; Nguyen et al., 2024). In connection with speaking skills, ChatGPT helps students overcome speaking anxiety by fostering their critical thinking in English speaking skills and increasing enjoyment (Cong-Lem et al., 2023), which is contradicted by Nguyen (2023). Moreover, Nguyen et al. (2024) and Tran and Tran (2023) concurred that ChatGPT can provide a stress-free learning environment, intensify learner motivation and decrease language anxiety, so it is described as a patient partner that is ideal for practicing language skills. Regarding learner autonomy, Nguyen and Tran (2023) and Nguyen (2024) claimed that ChatGPT serves as a valuable self-learning tool aiding learners to practice language skills and experiment with diverse phrases, grammar structures, and expressions without fear of being judged. Regarding creativity and critical thinking skills, ChatGPT can encourage learners to express their thoughts and creatively think outside the box (Hieu & Thao, 2024). ChatGPT provides a diversity of opinions and motivates students to consider different points of view (Nguyen, 2024). By communicating with ChatGPT and evaluating its answers, students develop skills necessary for their academic and professional lives (Nguyen, 2024). In contrast, Nguyen (2023) reports that ChatGPT tends to limit learners' creativity and autonomy as well as have a negative effect on their critical thinking skills. In addition, learners' creativity, independent thinking and language skills will be adversely affected (Cong-Lem et al., 2024; Hieu & Thao, 2024; Ho, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Thao et al., 2023; Yen et al., 2024).

4. Collaboration with Traditional Methods

AI tools like ChatGPT can collaborate with teachers' traditional teaching methods to improve educational outcomes (Yen et al., 2024), assessments (Nguyen, 2024; Nguyen & Tran, 2023), and test design (Nguyen, 2023). ChatGPT allows exploiting authentic linguistic resources which give opportunities for language learning (Cong-Lem et al., 2023; Nguyen, 2023; Tran & Tran, 2023). Thanks to this tool, teaching and learning strategies can be diversified (Tran & Tran, 2023).

The challenges of integrating ChatGPT with English education in Vietnam

Despite ChatGPT's benefits for English teaching and learning in Vietnam, considerable challenges remain.

1. Over-reliance and Academic Dishonesty

Learners tend to over-rely on ChatGPT for their language practice (Tran & Tran, 2023). This has its association with academic dishonesty due to students' dependence on ChatGPT with no effort; that is, issues of plagiarism and originality of students' work are unsolved (Cong-Lem et al., 2024; Nguyen, 2023; Nguyen & Tran, 2023; Nguyen et al., 2024).

2. Inconsistent Responses and Reliability Issues

Another concern is related to ChatGPT's inconsistent or irrelevant responses (Nguyen, 2024; Nguyen, 2024; Nguyen & Tran, 2023). The shortage of the exactness and suitability of ChatGPT's generated content negatively impacts its reliability as well as students' language abilities (Hieu & Thao, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Thao et al., 2023; Tran & Tran, 2023). Nonetheless, the results stemming from Cong-Lem et al. (2023), Ho (2024), Nguyen et al. (2024), and Thao et al. (2023) demonstrate that learners still improve their target language skills such as writing, speaking, or reading.

3. Lack of Training

There is no training or guidance on how to integrate ChatGPT into teachers' teaching practices effectively (Cong-Lem et al., 2024; Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Yen et al., 2024). Regarding the matter, Tran and Tran (2023) found that instructors face difficulties in using ChatGPT in classrooms and balancing it with traditional teaching methods. This disagrees with Nguyen (2023), Nguyen and Tran (2023), and Nguyen (2024) because their findings indicate that the tool supports teachers' work well (Nguyen, 2023; Nguyen, 2024; Nguyen & Tran, 2023; Yen et al., 2024).

4. Ethical and Privacy Concerns

It is urgent to increase awareness of potential hazards and limitations of the tool such as ethical and data privacy issues in educational settings (Nguyen, 2023; Nguyen, 2024; Nguyen, 2024; Nguyen & Tran, 2023; Thao et al., 2023; Yen et al., 2024).

Research question 2: What future trends and emerging research areas can be identified in the application of ChatGPT for English language teaching and learning?

After the use of the VOSviewer software, a bibliometric analysis indicates research trends in terms of utilizing ChatGPT for English education. The detailed analysis of the data discloses the cluster of keywords in the titles and abstracts from the research papers supplying emerging research areas and trends of the topic in the future.

1. Teacher Roles and Training

Instructors' role is considered fundamental in the digital environment where using AI like ChatGPT is needed for the advancement of educational practices and innovative pedagogical approaches (Nguyen et al., 2024). Furthermore, instructing learners to use ChatGPT properly is the teachers' mission in English language classrooms; therefore, there is a need for training teachers to acquire a detailed knowledge of this kind of technology (Cong-Lem et al., 2024; Hieu & Thao, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen et al., 2024).

2. Understanding Student Perspectives

It is important to understand students' attitudes and behaviors because teachers can support them well and provide more personalized learning experiences (Ho, 2024; Nguyen, 2023; Nguyen et al., 2024; Yen et al., 2024).

3. Collaboration Between AI and Human Educators

AI or ChatGPT cannot replace human teachers, so having a deeper understanding of teachers' perceptions towards the topic will help improve the quality of the AI-human collaboration in English education as well as their teaching practices from now on (Cong-Lem et al., 2024; Hieu & Thao, 2024; Ho, 2024; Nguyen, 2023; Nguyen, 2024; Nguyen, 2024; Nguyen & Tran, 2023; Nguyen et al., 2024; Yen et al., 2024).

4. Ethical Use and Data Privacy

The information ChatGPT delivers to learners is worth mentioning because what it offers is unlikely to be appropriate and accurate (Hieu & Thao, 2024; Nguyen, 2024; Thao et al., 2023). Discovering learners' habits of using ChatGPT for English language learning also aids educators in having effective teaching strategies (Ho, 2024).

Conclusion

The systematic review paper discussing 12 selected articles about the collaboration between ChatGPT and English education in Vietnam proved that the tool is promising and innovative for students' learning experiences and teachers' teaching practices.

The tool provides personalized learning environments where learners tend to get quick assistance, motivation, engagement, satisfaction, improved learning outcomes, simplified learning tasks, and interesting language practice through authentic resources and interactions, which contributes to their development of skills and abilities such as language skills, learner autonomy, creativity, and critical thinking skills. Besides, it helps instructors save their work,

time, and effort. Thanks to its aid, they can diversify strategies by combining them with traditional methods. However, ChatGPT also threatens to have negative influences on English language teaching and learning, namely learners' over-reliance on the tool, a lack of human interaction, academic dishonesty, and inaccurate or irrelevant answers. Additionally, having a strong knowledge of ChatGPT and how to integrate it with teaching practices properly needs to be considered for training activities.

Overall, learners can benefit from ChatGPT in terms of personalized learning, assistance, feedback, learner autonomy, engagement, reduced language anxiety, individualized language practice, learning outcomes, language translation, access to authentic resources, authentic language use, exposure to diverse language use, learning tasks, and development of skills and abilities, namely language skills, critical thinking skills, digital literacy skills, soft skills, and creativity. Furthermore, ChatGPT facilitates teachers' workload, assessment, and test design. These advantages contribute to diversified teaching and learning strategies and promising collaboration between the tool and human instructors in the future. Regardless of the benefits of ChatGPT, there remain numerous challenges that should be taken into account such as the quality and accuracy of AI-generated content, ethical and privacy issues, technical and resource limitations, cultural alignment challenges, academic dishonesty, over-reliance on the tool, reduced human interaction, and risk of promoting laziness, increasing the potential to replace instructors. Moreover, teachers struggle to balance ChatGPT with traditional teaching methods, and monitor and assess ChatGPT use, proving their lack of training.

There exist limitations in the review. The number of selected research papers is limited because inadequate studies regarding the specific use of ChatGPT in English teaching and learning in the context of Vietnam have been carried out compared to other countries, making it difficult for the review to represent the broader research landscape. It is suggested that there should be more studies chosen in future literature review articles. The contexts of selected articles included in the review vary from schools to universities, leading to different mindsets of students and teachers due to varying levels. However, the number of papers conducted between schools and universities is unequal, failing to represent the objectivity in the review. Reviews related to the topic ought to be done at a specific educational level in order to have a better understanding of its context. In addition, the limitations of the literature hamper comprehension of how ChatGPT can be effectively integrated into English education in Vietnam. The insufficient training on preventing plagiarism issues recommends a need for additional research to maintain academic integrity while using the tool (Ho, 2024). Teachers' unfamiliarity with ChatGPT, in spite of initial hands-on experiences, emphasizes a barrier to the adoption of the AI tool in English education (Nguyen, 2023). It is essential to conduct further research to develop training programs equipping instructors with adequate knowledge to make better use of the tool in their teaching practices, which will lead to extensive use of ChatGPT in EFL classrooms and enhance teaching and learning experiences. The limited number of learner and instructor participants in the existing literature prohibits generalizing the results to a wider setting of Vietnam (Nguyen et al., 2024; Thao et al., 2023; Yen et al., 2024). This requires a broader and more representative sample in the next research to comprehend ChatGPT's effects on various regions and contexts. Additionally, focusing on one qualitative method (Tran & Tran, 2023; Thao et al., 2023; Hieu & Thao, 2024; Yen et al., 2024; Nguyen, 2024; Nguyen &

Tran, 2023) or one quantitative method (Nguyen, 2024; Cong-Lem et al., 2024) instead of mixed-method approaches limits the depth of perspectives on the utilization of ChatGPT in EFL education. A mix of qualitative and quantitative methods in future studies can supply a comprehensive and detailed analysis of the tool's advantages, disadvantages, and effectiveness. Besides, conducting research in a specific context may limit the applicability to distinct cultural and educational settings (Hieu & Thao, 2024). Papers in the future should include a wide variety of settings to examine the generalizability of the findings, assisting in exploring the impacts of heterogeneous cultural and educational environments on the effectiveness of ChatGPT.

Acknowledgments

References

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. (2023). Impact of ChatGPT on learning motivation: teachers and students' voices. *Journal of English Studies in Arabia Felix*, *2*(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Aljanabi, M. (2023). ChatGPT: Future Directions and Open Possibilities. *Mesopotamian Journal of Cybersecurity*, 16-17. https://doi.org/10.58496/MJCS/2023/003

Aromataris, E., & Munn, Z. (2020). *JBI Manual for Evidence Synthesis*. Joanna Briggs Institute: Adelaide, Australia.

Averis, A., & Pearson, A. (2003). Filling the gaps: identifying nursing research priorities through the analysis of completed systematic reviews. *JBI Evidence Implementation*, *1*(3), 49-126. https://doi.org/10.1046/j.1479-6988.2003.00003.x

Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology*, *24*(1), 28-47. https://doi.org/10.56297/FSYB3031/MXNB7567

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, *61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Egger, M., Smith, G. D., & O'Rourke, K. (2001). Introduction: rationale, potentials, and promise of systematic reviews. *Systematic reviews in health care: meta-analysis in context*, 1-19. https://doi.org/10.1002/9780470693926.ch1

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR medical education*, *9*(1). https://doi.org/10.2196/45312

Gough, D., Thomas, J., & Oliver, S. (2017). *An introduction to systematic reviews*. London: SAGE Publications Ltd. Retrieved from http://digital.casalini.it/9781473968219

Hieu, H. H., & Thao, L. T. (2024). Exploring the Impact of AI in Language Education: Vietnamese EFL Teachers' Views on Using ChatGPT for Fairy Tale Retelling Tasks. *International Journal of Learning, Teaching and Educational Research*, *23*(3), 486-503. https://doi.org/10.26803/ijlter.23.3.24

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

Hoang, N. T., Duong, N. H., & Le, D. H. (2023). Exploring Chatbot AI in improving vocational students' English pronunciation. *AsiaCALL Online Journal*, *14*(2), 140-155. https://doi.org/10.54855/acoj.231429

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation, 5*(1), 37–45. https://doi.org/10.61414/jeti.v5i1.103

Iskender, A. (2023). Holy or unholy? Interview with open AI's ChatGPT. *European Journal of Tourism Research, 34*, 3414. https://doi.org/10.54055/ejtr.v34i.3169

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the royal society of medicine*, *96*(3), 118-121. https://doi.org/10.1177/014107680309600304

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *TESL-EJ*, *27*(3), 1-19. https://doi.org/10.55593/ej.27107int

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Lozano, A., & Fontao, C. B. (2023). Is the education system prepared for the irruption of artificial intelligence? A study on the perceptions of students of primary education degree from a dual perspective: Current pupils and future teachers. *Education Sciences, 13*(7), 733. https://doi.org/10.3390/educsci13070733

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology, 50*(6), 2824–2838. https://doi.org/10.1111/bjet.12861

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024, February). The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. In *Frontiers in Education* (Vol. 9, p. 1328769). https://doi.org/10.3389/feduc.2024.1328769

Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in

enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Education and Information Technologies, 29*, 3195–3217. https://doi.org/10.1007/s10639-023-11917-z

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, *12*(8), 153. https://doi.org/10.3390/computers12080153

Mulyawati, I. B., & Ramadhan, D. F. (2021). Bibliometric and visualized analysis of scientific publications on geotechnics fields. *ASEAN Journal of Science and Engineering Education*, *1*(1), 37-46.

Nandiyanto, A. B. D., & Al Husaeni, D. F. (2021). A bibliometric analysis of materials research in Indonesian journal using VOSviewer. *Journal of Engineering Research, 9*, 1–16. https://doi.org/10.36909/jer.ASSEEE.16037

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. C. (2024). University Teachers' Perceptions of Using ChatGPT in Language Teaching and Assessment. *Proceedings of the AsiaCALL International Conference*, *4*, 116–128. https://doi.org/10.54855/paic.2349

Nguyen, T. H. B., & Tran, T. D. H. (2023). Exploring the Efficacy of ChatGPT in Language Teaching. *AsiaCALL Online Journal, 14*(2), 156-167. DOI: https://doi.org/10.54855/acoj.2314210

Nguyen, T. P. T. (2023). The Application of ChatGPT in Language Test Design – The What and How. *Proceedings of the AsiaCALL International Conference*, *4*, 104–115. https://doi.org/10.54855/paic.2348

Nguyen, T. Q. Y. (2024). Unraveling the Potential of ChatGPT: Investigating the Efficacy of Reading Text Adaptation. *Proceedings of the AsiaCALL International Conference*, *4*, 159–169. https://doi.org/10.54855/paic.23412

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. DOI: https://doi.org/10.54855/ijli.23231

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Pham, M. T., Luu, T. T. U., Mai, T. H. U., Thai, T. T. T., & Ngo, T. C. T. (2022). EFL Students' Challenges of Online Courses at Van Lang University during the COVID-19 Pandemic.

International Journal of TESOL & Education, 2(2), 1-26. DOI: https://doi.org/10.54855/ijte.22221

Pham, M. T., Nguyen, T. T. V., & Cao, T. X. T. (2023). Vietnamese Students' Perceptions of Utilizing Social Media to Enhance English Vocabulary: A Case Study at Van Lang University. *International Journal of TESOL & Education, 3*(1), 79-111. DOI: https://doi.org/10.54855/ijte.23316

Pham, V. P. H., & Le, A. Q. (2024). ChatGPT in Language Learning: Perspectives from Vietnamese Students in Vietnam and the USA. *International Journal of Language Instruction*, *3*(2), 59–72. https://doi.org/10.54855/ijli.24325

Pollock, A., & Berge, E. (2018). How to do a systematic review. *International Journal of Stroke*, *13*(2), 138-156. https://doi.org/10.1177/1747493017743796

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of development effectiveness*, *4*(3), 409-429. https://doi.org/10.1080/19439342.2012.710641

Spratt, C., Walker, R., and Robinson, B. (2004). *Mixed research methods: Practitioner research and evaluation skills training in open and distance learning*. Burnaby, BC: Commonwealth of Learning. Retrieved from https://oasis.col.org/colserver/api/core/bitstreams/a0f649ca-9764-438a-b10b-84638b956ebf/content

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, *20*, 1-17. https://doi.org/10.1186/s41239-023-00404-7

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31–40. https://doi.org/10.37074/jalt.2023.6.1.17

Thao, L. T., Hieu, H. H., & Thuy, P. T. (2023). Exploring the Impacts of ChatGPT in EFL Writing: Student Perceptions of Opportunities and Challenges in Vietnamese Higher Education. *Kognisi: Jurnal Ilmu Keguruan*, *1*(2), 107-124.

Tran, T. N., & Tran, H. P. (2023). Exploring the Role of ChatGPT in Developing Critical Digital Literacies in Language Learning: A Qualitative Study. *Proceedings of the AsiaCALL International Conference*, *4*, 1–17. https://doi.org/10.54855/paic.2341

Tri, P. M., Van, N. T. T., & Tu, C. T. X. (2023, April). EFL Students' Perceptions and Challenges of Computer-Mediated Collaborative Writing in Academic Writing Courses at Van Lang University. In *19th International Conference of the Asia Association of Computer-Assisted Language Learning (AsiaCALL 2022)* (pp. 34-54). Atlantis Press. https://doi.org/10.2991/978-2-38476-042-8_5

Tricco, A. C., Tetzlaff, J., & Moher, D. (2011). The art and science of knowledge synthesis. *Journal of clinical epidemiology*, *64*(1), 11-20.

https://doi.org/10.1016/j.jclinepi.2009.11.007

Yang, H., & Kyun, S. (2022). The current research trend of artificial intelligence in language learning: A systematic empirical literature review from an activity theory perspective. *Australasian Journal of Educational Technology*, *38*(5), 180-210. https://doi.org/10.14742/ajet.7492

Yen, P. H., Thu, H. T. A., Thi, N. A., Tra, N. H., & Thuy, P. T. (2024, May). University Teachers' Perceptions on the Integration of ChatGPT in Language Education Assessment: Challenges, Benefits, and Ethical Considerations. In *20th International Conference of the Asia Association of Computer-Assisted Language Learning (AsiaCALL 2023)* (pp. 7-21). Atlantis Press. https://doi.org/10.2991/978-94-6463-396-2_2

Zhai, X. (2022). *ChatGPT user experience: Implications for education*. Retrieved from http://dx.doi.org/10.2139/ssrn.4312418

Biodata

Pham Manh Tri received his bachelor's degree in English Language from Van Lang University and is currently pursuing a master's degree at the same institution. He works as an English teacher for children, teenagers, and adults and a Vietnamese teacher for English speakers. He has presented his research papers at numerous international conferences and published several articles in international peer-reviewed journals. Additionally, he serves as a peer reviewer for an international journal. His research interests include online education, language skills, TESOL, and technology integration in language education.

Cao Thi Xuan Tu works as a lecturer at the Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam. She has taught English for more than eight years and has experience in teaching English to a variety of learners ranging from young learners to undergraduate students. She evaluates and designs teaching materials and courses for English-majored students, such as Phonetics and Phonology, Writing, Listening, and Interpretation. She is also a reviewer of the International Journal of TESOL & Education. She is implementing studies on Task-based Language Teaching (TBLT), computer-mediated tasks, and collaborative writing.

Augmented Reality in English Language Teaching: A Literature Review on Catering to Diverse Learning Styles

Ngo Nguyen Thien Duyen[1], Vo Trong Nghia[1*]

[1] Ho Chi Minh City University of Economics and Finance, Vietnam
[*]Corresponding author's email: nghiavt@uef.edu.vn
[*] https://orcid.org/0009-0004-0560-3967
https://doi.org/10.54855/ijte.25514

ABSTRACT

In contemporary educational contexts, the application of Augmented Reality (AR) in English Language Teaching (ELT) has gained significant attention. Thus, the aim of this literature review is to investigate the benefits of AR in accommodating diverse learning styles and individual learner needs within ELT. AR's multisensory features cater effectively to visual, auditory, and kinesthetic learners, enhancing engagement and motivation while promoting inclusivity. For visual learners, AR offers interactive visuals and 3D models; auditory learners benefit from pronunciation guides and immersive dialogues, while kinesthetic learners engage through hands-on interaction with virtual elements. The analysis, grounded in the Universal Design for Learning (UDL) framework, highlights AR's potential in providing equitable learning opportunities. This review synthesizes current research and intends to offer insights to educators and developers who want to utilize Augmented Reality to design language learning experiences that are effective, engaging, successful, and inclusive. It synthesizes current research findings to achieve this goal.

Introduction

The term *Learning styles* are defined as the naturally varying tendencies of people in perceiving, processing, and retaining information (Dunn & Dunn, 1993). Since then, several distinguished descriptive models have emerged to visualize the spectrum of learners' comprehension approaches. The patterns of learning styles can be classified into a limited number of groups. Students acquire knowledge by visual, auditory, or kinesthetic means, with varied levels of efficacy. It is crucial to acknowledge the diverse range of students in the English classroom in order to implement suitable strategies to support their individual needs and facilitate their success, as it can help develop students' learning and motivation.

As such, the current surge in promoting inclusivity in education, specifically in designing an optimal learning environment that caters to the requirements of all learners, is closely connected

to addressing the problem of intrinsic learning styles. When striving for inclusion, educators must guarantee equal access to knowledge and learning opportunities for all students, regardless of their backgrounds, skills, or preferences in the classroom (UNESCO, 2005). To achieve that objective, it is necessary to have instructions and materials that can be easily adjusted to meet the specific requirements of the learners (Tomlinson, 2014).

Technology has become a transformative force in addressing these diverse learning needs in modern educational contexts. Digital tools and platforms allow educators to design interactive, personalized, and engaging experiences that cater to individual learner preferences. The integration of technology in ELT enhances accessibility, promotes inclusivity, and supports language acquisition through immersive and interactive approaches (Chen, 2020; Kukulska-Hulme & Shield, 2008). In response to said demands, Augmented Reality (AR) technology arises as an auspicious instrument to improve learning by integrating digital content into our physical environment (Dunleavy et al., 2009). AR generates interactive and immersive experiences for users as it fuses virtual objects, information, or multimedia with reality (Azuma, 1997). It thus provides a distinctive approach to accommodating a variety of learning styles when the users perceive the real world with supplemented visuals and audio, in contrast to Virtual Reality (VR), which simulates entirely new environments (Milgram & Kishino, 1994).

Several studies have looked at the potential of AR in the classroom to improve language instruction and student performance (Klopfer et al., 2002; Wu et al., 2013). Using AR, field educators may immerse students in a more realistic setting where they can practice using real language, which is sure to pique their interest (Cheng & Tsai, 2013). Nevertheless, a number of research studies (Economides et al., 2020; Li & Wong, 2021; Liu et al., 2023b) fail to address higher-order cognitive abilities such as reading comprehension, writing ability, and intercultural competency in favor of assessing lower-order language skills, such as vocabulary development.

Also, upon most parts of modern life being incorporated with technology that is media-incentivized, consensus believes it is not viable to promote a unitary product design philosophy, as the ways men acquire assistance naturally vary (Abascal & Nicolle, 2005; Ladau, 2021). Against such a backdrop, AR's interactive and multimodal experiences have the potential not only to avoid this but also to greatly improve the inclusion of users. However, up until this point, ELT research has either focused on very narrow AR applications or used very small sample sizes (e.g., groups of college students) to investigate the effects of AR, leaving a vast array of situations unexplored. Possible applications include bringing together students of wildly varying ages, ethnicities, and socioeconomic statuses, as well as those from quite diverse classrooms (Pachler et al., 2010; Yen et al., 2013). Thus, a gap remains in understanding the use of AR for inclusive learning.

Therefore, this literature review aims to fill the void by investigating the existing research on integrating AR in ELT; in particular, the study focuses on how AR can help promote inclusivity in classroom practices. The current study hopes to provide a comprehensive view of the field by applying the theoretical frameworks of Universal Learning Design and examining previous studies' empirical results and recommended practices. Besides, studies related to technology integration should regularly be updated due to the rapid advancement in the field; therefore, this paper aims to present the updates from recent studies. Overall, this literature review paper hopes to provide education stakeholders with a more nuanced understanding of using AR effectively to create meaningful language learning experiences.

Literature Review

Definition and Conceptualization of AR in the Context of Language Learning

The integration of digital technology in ELT has made the revolution of well-tested and used English Language Teaching (ELT) methodologies possible. The learners' engagement, accessibility, and personalization have been enhanced by technologies and educational tools (Kukulska-Hulme & Shield, 2008). However, AR has become popular among these technologies due to its immersive and interactive learning environments that address different learning needs (Dunleavy et al., 2009). AR adds virtual objects, sounds, and multimedia into physical locations to bring the real world to life (Azuma, 1997; Milgram & Kishino, 1994). This contextual blending means learners have chances to practice language skills in real situations, through which motivation and learning anxiety (Chang et al., 2020; Huang et al., 2021) could be reduced.

Klopfer's (2008) current conception of AR can be paraphrased as the blending of the real world with additional information relevant to the real world. It does this by generating sounds and projecting images into the user's field of vision. Sannikov et al. (2015) and Liu et al. (2023a) later affirm AR's evolvement, with games of educational values on mobile devices becoming its fertile ground.

In teaching and learning languages, stacking layers of virtual elements over reality has helped engage and motivate learners (Wedyan et al., 2022; Min & Yu, 2023; Liu et al., 2023a), although how it does this remains vague. An indication is Liu et al. (2023b)'s report of the enjoyment and fulfillment of students being facilitated by AR, which is concluded from self-reports that may be under social desirability bias. Thus, approaches with a higher degree of objectivity may be required to investigate and solidify future findings, namely methods like physiological indicators of engagement.

Studies on AR in ELT emphasize its potential to enhance various language skills, including vocabulary acquisition, listening comprehension, and speaking proficiency (Chen, 2020; Yulian et al., 2022). For instance, AR's visual and auditory elements help learners comprehend abstract concepts more effectively by providing contextual and multimodal inputs (Liu et al., 2023a). Kinesthetic learners benefit from AR's interactive features, such as manipulating virtual objects and participating in gamified learning tasks (Iqbal & Campbell, 2021).

Literature offers more evidence of AR's ability to enable authentic learning. This is seen with AR used via mobile devices (Lee & Park, 2020; Pellas et al., 2019) as the added visuals, texts, and sounds onto reality embed contextual information into learners' daily lives. Interacting with those elements gives users a deeper situated learning of language (Chang et al., 2020; Liu et al., 2023a). What is also reduced is learning anxiety. Thanks in part to AR's playfulness and immersiveness, these elements invite learners to explore, actively participate in the scaffolding activities, and take risk-free initiatives to learn (Huang et al., 2021). That said, stand-alone reports from Chang et al. (2020) and Liu et al. (2023a) still do not sufficiently contribute to AR-based situated learning, and the causal link between such capacity and the improvement of students' language proficiency remains to be explored in future research using randomized controlled trials or other more rigorous method designs.

Language skills are the next dimension in AR's assistive capacity. Empirical tests of AR integration point to improved vocabulary (Huang et al., 2021), listening and speaking (Chang et al., 2020) as well as reading (Yulian et al., 2022; Şimşek & Direkçi, 2023) and academic writing (Lin et al., 2020). That said, the applied scopes of these findings are partially undermined when many of them are reached through investigating specific AR applications or

limited learner populations. To give evidence, Huang et al. (2021)'s experimental group was a small sample size of university students learning vocabulary with AR, while Chang et al. (2020)'s setting was set in a junior high school, where learners were taught and examined in their listening and speaking skills only. Results of the applicability of AR in ELT need to be equally evident in alternative age groups (Bistaman et al., 2018), proficiency levels, and other educational situations.

UDL has a conceptual basis to contextualise AR as a means to foster inclusivity in ELT. In providing different means of representation, engagement, and expression, AR allows the learner who prefers and is able to represent, engage with, or express themselves in one of these ways (Hall et al., 2012). For example, AR's 3D animations and diagrams are great for those who are visual learners, audio guides and dialogues help auditory learners and environments that allow kinesthetic learners to promote physical interaction and exploration (Iqbal & Campbell, 2021; Chen, 2020; Wu, 2019).

However, there are challenges to implementing AR. Access to AR technology is limited, plus teachers and teachers in general are not well trained, and there is no culturally responsive content in the curriculum (Lee & Park, 2020; Manna, 2023). In addition to this, it is possible that the novelty effect experienced when using AR may diminish over time, and attending to sustaining efforts to integrate AR meaningfully into the curriculum (Deterding et al., 2011).

Although the benefits of engagement and engagement of AR for language learners are well documented, deeper investigations of AR directives included in AR experiences that maintain engagement and inclusion of language learners are needed. For example, the causal relationship between AR's immersive features and long-term language proficiency has not been researched (Marrahí-Gómez & Belda-Medina, 2022). Furthermore, research especially related to the employment of AR for different demographies and dissertations about the longitudinal AR impacts are very limited (Pachler et al., 2010; Fombona et al., 2017).

Finally, in sum, AR is beneficial to ELT due to its ability to improve levels of engagement, motivation, and skills, but more work still needs to be done to conceptualise AR as an inclusive tool to support the diversity of learners' needs.

Theoretical Foundations - Universal Design for Learning

In response to *learning styles,* the Universal Design for Learning (UDL) stands as an inclusive approach that assists all students in varied modern classrooms. It is applied when research, development, technology, and educational practice are directed toward a course, helping it strategically anticipate all possible requirements of students and then extending the planning process to include the whole scope of a classroom. Rose and Meyer (2002) see it as an attempt to escape the discrepancy between the growing diversity of the students and the standardized curriculum that would not lead to the desired academic improvements.

The UDL framework is demonstrated as growing from understanding brain development, learning, and digital media (Rose & Meyer, 2002) and comprises three principles. These principles emphasize the importance of offering many options, specifically for representation, action, and expression, as well as engagement (Meyer et al., 2014).

The first principle emphasizes the importance of presenting information and knowledge through various methods (e.g., representation) to enable students to acquire, process, and integrate materials effectively (Meyer et al., 2014). This approach is based on the understanding that students may face disparities in their comprehension processes due to auditory, visual, linguistic, cultural, or cognitive limitations. As a result, no single method of representation suits

all learners, making the availability of choices an essential element in inclusive education (CAST, 2018).

Equally significant is the second principle, which advocates for a diverse array of instructional strategies and thus allows students multiple ways to demonstrate their understanding of the material (Hall et al., 2012). In its rationale, certain pupils would sufficiently articulate their knowledge in written form but struggle with verbal expression, while others encounter the opposite challenge. These variations often stem from differences in physical capabilities, language proficiency, or distinct learning strategies. Since a fixed, imposed set of expected demonstration approaches will not be ideal for all learners, offering a broad spectrum of possible actions ensures that educational practices accommodate diverse needs and foster equitable opportunities for student success.

The third concept, "provide multiple means of engagement," refers to the need to offer a variety of choices to enhance student motivation (Hall et al., 2012) and attention during learning (Meyer et al., 2014). The fundamental premise of this notion caters to learners' emotional circumstances in the learning process, which several sources, including neurological factors, cultural influences, personal significance, subjectivity, and knowledge background, may influence. Certain learners have a strong inclination towards novelty and spontaneity, while others harbor a dislike for new experiences and tend to be apprehensive, preferring a more predictable routine. Some learners want to work alone, while others enjoy collaborating in groups. It is necessary to offer many alternatives to encourage and boost the interest of students since there is no one method that can motivate or improve the engagement of all learners (CAST, 2018).

From the principles of UDL, integrating AR into ELT promises to improve inclusivity, making it more accessible and engaging for all learners (Rose et al., 2018). In summary, the evidence points to AR capabilities aligning with UDL principles. AR ushers in more inclusive and learner-centered ELT by offering representation, engagement, action, and expression channels. It also pushes motivation, participation, and learners' ownership of knowledge (Meyer et al., 2014).

Previous Studies & Research Gap

As stated earlier, the documented benefits of AR for learner engagement, motivation, and practical language skills (Huang et al., 2021; Chang et al., 2020) still raise questions about how the tool might be leveraged to its full potential for varied learning preferences and thereby promote inclusivity in the ELT landscape. Such inclusiveness needs to extend to wider age groups, cultures, and educational settings (Pachler et al., 2010; Yen et al., 2013).

While emerging research has begun to recognize AR's potential for inclusive learning, a scarcity in more holistic investigations remains, namely in terms of implementation challenges, technological accessibility, teacher training, and pedagogical design (Kukulska-Hulme & Shield, 2008; Manna, 2023; Qiu et al., 2023). This is hindering evidence-based guidelines and good practices to integrate AR effectively into diverse ELT contexts.

Based on this observation, the current study is diving deeper to shed light on AR's full capacity in ELT. We hope that future research areas can be identified that would maximize the technology's potential of enabling inclusive and effective language mastery.

Research Questions

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

1. How does AR cater to diverse learning styles in the context of ELT?

2. What strategies can be employed to design AR experiences that are inclusive of individual learner needs?

Methods

Design of the Study

A systematic literature review serves as the basis for this study to analyze how Augmented Reality (AR) incorporates different learning styles and individual learner interests and thus enhances the process of Inclusive English Language Teaching (ELT). A systematic review is particularly suitable for synthesizing existing evidence and providing comprehensive insights on a particular topic. This method involves the systematic analysis of peer-reviewed studies so as to ensure the inclusion of high-quality research and facilitate pattern, gap, and trend identification in the literature (Popay et al., 2006).

The essence of this methodology is based on the exact objectives of the study. The systematic review integrates more than one source of finding to provide answers to broad questions on how AR fits with Universal Design for Learning (UDL) principles and the readiness of the UDL to support inclusivity in various educational environments. Moreover, this approach to researching AR technology and its potential in the ELT field is appropriate due to the fast development of AR technology and timely coverage of its applications.

Data collection & analysis

This stage begins as authors sift through various academic databases extensively. Through Google Scholar, ERIC, JSTOR, and the Web of Science, various combinations of keywords and terms related to the theme of interest are selected: "augmented reality," "English language teaching," "learning styles," "inclusive design," "personalized learning," and "language acquisition."

Studies were further filtered with specific criteria, as shown in Table 1.

After the screening process comes thematic analysis, which requires in-depth reading to extract recurring patterns and insights from reports that align with this literature's investigative goals. The information was then categorized so that the following core themes emerge:

- AR's capacity towards Learning Styles: authors discovered findings into how AR caters to different preferences (e.g., visual, auditory, kinesthetic) (Chen & Tsai, 2013; Huang et al., 2020). These eventually indicate that the tool can be particularly effective for learners long acquainted with multimodal inputs and experiences (Wu et al., 2013).

- AR's design strategies for inclusivity: The literature enumerates the considerations that went into building AR experiences that ensure its accessible, personalized, flexible, and culturally responsive uses, friendly to the diverse types of learners (Ke & Hsu, 2015; Martín-Gutiérrez et al., 2015). They include adjustable difficulties, optional interaction modalities, and cultural sensitivity in content and design (Zhao et al., 2018).

After this rigorous screening process, the existing literature is analyzed comprehensively. Key

insights are revealed, along with evidence-based practices and areas where further research can be fruitful. The authors believe that our work has given us a deeper understanding of how AR can transform language learning.

Table 1.

Inclusion and Exclusion Criteria for Study Selection

Criteria	Inclusion	Exclusion
Focus	Studies explicitly address the use of AR in ELT and its relation to learner needs.	Studies do not explicitly address the use of AR in ELT and its relation to learner needs.
Publication Type	Peer-reviewed journal articles, book chapters, conference proceedings, and credible reports from recognized organizations.	Non-peer-reviewed articles, opinion pieces, blog posts.
Publication Date	Published within the last ten years (2014-2024)	Studies published before 2014
Language	Published in English.	Published in languages other than English.
Methodology Quality	Empirical studies with clear methodology and significant findings; rigorous qualitative, quantitative, or mixed methods studies.	Studies with unclear methodology, low-quality, or inconclusive findings.
Relevance	Directly relevant to research questions/themes: learning styles, learner needs, inclusivity, and strategies.	Indirectly related or irrelevant to the core themes of the review.

Findings

The literature from this review gives a fulfilled response to the research questions. It signifies AR's potential and how its designs can elevate individuals of varied learning orientations.

AR and Diverse Learning Styles

AR's multisensory feature deeply resonates with learners, and data support this claim. Visual, auditory, and kinesthetic incorporations within AR represent versatile learning modalities, making it inclusive for language learners of varying preferences.

Visual learners who favor visual information may find AR images beneficial. Several studies conclude that comprehension and retention are better when AR's 3D animations make abstract concepts tangible and interactive (Kalyuga, 2009; Chen, 2020). For instance, vocabulary lessons are brought to life as AR allows learners to manipulate virtual objects associated with the words at hand. This approach makes learning more engaging and helps learners establish stronger connections between words and their meanings.

Auditory learners who thrive through listening and verbal communication may also find great help in AR's audio features, namely pronunciation guides, dialogues with real-time feedback, and interactive conversations with virtual characters. Combined, these facilitate realistic yet repeated opportunities for learners to practice their listening and speaking skills (Pasfield-

Neofitou, 2014; Wu, 2019), perhaps surpassing the restraints of real-life conversations.

Iqbal and Campbell (2021), corroborating with the findings of Huang et al. (2021), believe kinesthetic learners, who learn best through hands-on experiences, are enabled to move and interact extensively thanks to AR environments. What caters to them are AR games with challenges such as letting learners search for items, map out the simulated space, or find answers to language-based puzzles.

AR also accommodates those with reading/writing or global/analytic preferences. Their favored text-based information is a simple addition for AR alongside visual and auditory content, reinforcing their understanding through different modalities, like writing exercises and quizzes (Lin et al., 2020).

Kalyuga (2009) adds that AR offers choices when presenting information. Global learners want to show the big picture before focusing on details, so AR's templates of overarching context and visual overviews help them the most. In adjacent, analytic learners tend to break down information into smaller components, which augmented interactive features enable them to do structurally. This is done with virtual tasks that require step-by-step analyses of information and attention to specific details.

Inclusive AR Design Strategies

The second investigative question that gives attention to inclusive AR design strategies is well addressed in the literature, most of which build their theories upon the Universal Design for Learning (UDL) framework. The literature believes AR can effectively implement the UDL principles of emphasizing multiple means of representation, action, expression, and engagement (CAST, 2018), thus benefiting teachers, administrators, and institutions in developing pedagogical strategies. That said, a more thorough examination recognizes potential strengths as well as limitations in these claims.

Firstly, the praises towards AR for its multiple means of representation, while tangible, may be overstated. The AR system can offer various visual elements such as 3D models, animations, and diagrams, which are claimed to make complex concepts sufficiently illustrious for visual learners (Chen, 2020). Similarly, auditory learners seemingly gain from audio narration, pronunciation guides, and interactive dialogues (Wu, 2019), and textual inputs adhere to those who prefer reading and writing (Lin et al., 2020). Moreover, culturally considerate designs with relevant imagery and perspectives are believed to enable more seamless learning (Lee & Park, 2020). However, the effectiveness of these multimodal approaches in genuinely enhancing learning outcomes across diverse learner profiles is often concluded upon anecdotal evidence, and empirical findings remain under-researched (Beetham & Sharpe, 2019; Meyer et al., 2014; Puentedura, 2013).

Secondly, AR is lauded for enabling flexibility for learners to interact and varied means to express their understanding, supposedly empowering multiple communicative styles (Huang et al., 2021). An example is when AR language learning apps may give different choices for vocabulary practices: saying words out loud, writing them down, or manipulating virtual items. However, such flexibility might not automatically translate to better learning outcomes. The extent to which these varied formats genuinely accommodate individual learning differences without overwhelming learners or diluting the learning focus requires further scrutiny (Hattie, 2009; Kirschner et al., 2006).

Thirdly, AR's immersive and interactive nature is often cited as inherently fostering engagement. While gamification elements like points, badges, and leaderboards (Liu et al., 2023a) are designed to motivate learners, the long-term effectiveness of such engagement strategies is

questionable. The novelty of AR may wear off, and reliance on gamification can lead to superficial engagement rather than deep, meaningful learning (Deterding et al., 2011; Nicholson, 2015). Personalized feedback and exploration of personal interests are posited to enhance ownership and autonomy (Lee & Park, 2020), yet these benefits are contingent on the quality and relevance of the feedback and the genuine alignment of content with learners' interests (Hattie & Timperley, 2007; Black & Wiliam, 1998).

Personalization in learning and teaching is heralded as a crucial benefit of AR. Differences in learners' needs and preferences can ostensibly be addressed through choices in interaction modes and activities (Liu et al., 2023b). However, implementing such personalization while ensuring effectiveness and practicality in the confines of the course is a complex, if not relatively infeasible, endeavor. To explain, it is essential that accessibility to features like captions, alternative input methods, and clear navigation (Wedyan et al., 2022) must be granted in a balanced manner, yet these solutions are rarely implemented with such calibrations in mind, potentially leaving AR's inclusiveness less fulfilling for some learners than others (Burgstahler, 2015; Seale, 2013).

In summary, despite the literature highlighting the potential of inclusive design in ushering in effective AR experiences for language learning, such a claim warrants a critical perspective. Adhering to UDL principles, personalizing content, ensuring accessibility, and incorporating flexibility and cultural responsiveness are promising strategies. However, the actual efficacy of AR in fostering genuinely inclusive and equitable language learning environments remains to be conclusively demonstrated. More rigorous, longitudinal research is needed to substantiate the optimistic claims made about AR's impact on education (Aguayo et al., 2017; Lin & Lan, 2015).

Discussion

AR and Learning Styles

This review contributes to earlier investigations of AR in English Language Teaching (ELT). Those findings assert its prospective enhancement of learners' capacity to engage in, to be incentivized by, and to become proficient in certain language skills (Economides et al., 2020; Chang et al., 2020; Garzón & Acevedoet, 2019; Cai et al., 2022; Liao et al., 2024; Wedyan et al., 2022). This review, on the other hand, broadens and exhausts the available reports, discerning AR's accumulated benefits for myriad learning styles thanks to its emphasis on inclusive design. The review does so by synthesizing overlooked evidence on the diverse needs of learners that AR designs can cater to in terms of personalization, accessibility, and cultural responsiveness. It unequivocally demonstrates the promises AR has towards ELT, that is, making the field more inclusive through adhering to various learning preferences.

AR's multisensory approach resonates with different ways students learn. 3D animations are useful to visual learners, as they provide better comprehension and recall (Kalyuga, 2009; Chen, 2020). Talking Aupair serves as a pronunciation guide and an interactive dialogue for auditory learners, thereby giving them sharpened listening and speaking proficiency (Pasfield-Neofitou, 2014; Wu, 2019). Huang et al. (2021) state that 'Kinesthetic learners tend to learn more effectively through hands-on interaction with virtual objects and environments. Furthermore, AR supports various ways of information and interaction from people with reading and writing and global and analytic preferences (Lin et al., 2020). These findings show that AR provides a more personalized and effective learning experience for each student because AR is an offer offering AR to each student (Akçayır & Akçayır, 2017; Bacca et al., 2014).

Inclusive Design Strategies

The review also makes reference to AR's appropriateness with the UDL based on a framework that emphasizes the creation of inclusive learning environments that can serve all learners' needs (CAST, 2018). By following the norms of the UDL personalization principle, we analyze the content elements to support learners' preferences to implement and customize the content difficulty, interaction modes, and activities per the individual requirements (Liu et al., 2023a). The ability to optimize interactions and ensure that all learners participate is very important. Moreover, FLP and cultural sensitivity are incorporated into AR, to avoid the learners' nonwestern orientation (Huang et al., 2021; Lee & Park, 2020). To help make AR accessible for physically challenged learners, also initiatives have been taken. Moreover, they include caption providing, alternative input methods providing, and user-friendly navigation features (Wedyan et al., 2022; López Belmonte et al., 2019). They also make it only more likely that AR will be a means of enabling equitable access to learning opportunities in a variety of educational settings.

Additional Emerging Themes

Beyond addressing the research questions, several additional themes emerged from the analysis. One prominent theme is the growing role of AR in fostering intercultural competence. Liu et al. (2023b) found that AR-based instruction led to such an affordance being better developed than under traditional teaching. This AR's ability to promote understanding and communication across contexts is worth putting in more effort to seek further insights. A stronger grasp is equally needed for the long-term effects of AR on language outcomes and its mix with broader pedagogical philosophies. Promising evidence for AR's short-term benefits has been provided, but longitudinal studies to assess its sustained impacts are scarce (Fombona et al., 2017). Additionally, it remains obscure how AR can be effectively integrated with other technologies and instructional methods to further the outcome of comprehensive and holistic language learning (Dunleav et al., 2009). Another significant factor influencing the success of AR implementation is learners' technology self-efficacy. As highlighted by Do et al. (2024), learners with higher self-efficacy in technology are more likely to engage effectively with digital tools. Addressing this aspect through training and supportive environments can mitigate barriers to AR adoption and ensure more equitable learning outcomes.

Consequently, the research findings show that AR effectively enhances more favorable and efficient ELT. The above aspects further substantiate the claims on how AR enables learners of varying attributes to meet their language learning needs through learning modalities, personalization, and UDL principles. This type of empowering is particularly crucial in ELT, primarily due to the fact that students' diversity levels are frequently high, and specific traditional approaches are insufficient to address their requirements. However, the successful implementation due to the necessary balancing of the pedagogical affordances or design, its technological availability, and the preparation of teachers (Belda-Medina & Calvo-Ferrer, 2022; Li & Wong, 2021; Qiu et al., 2023). Thus, future studies should focus on exploring the learning potential of AR as well as the practices of utilizing this tool in educational settings for teachers and heads of educational institutions that would help avoid ineffective and unequal use of the tool in different contexts for the ELT (Ibáñez & Delgado-Kloos, 2018).

Pedagogical Implications

Substantial attention has been placed on applying AR as a tool in contemporary language learning, corroborating findings from previous reviews (Bower et al., 2014; Akçayır & Akçayır, 2017). This literature synthesizes reported benefits of recent AR in accommodating diverse

learner needs within the context, those that are gained with the utilization of the framework of UDL.

Given the proven effectiveness of AR in our findings, it can be indicated that the trend towards their adoption in classrooms is imminent, and such is an appropriate tendency to enhance learning outcomes. To accommodate this shift, instructional materials should be made more compatible with these technologies. With various pedagogical approaches implemented, such as self-directed learning and task-based learning, the designs and developments of AR can help to further integrate itself into the curriculum. For instance, Hsu (2019) found that students experienced higher engagement in self-directed AR activities, suggesting that self-control in terms of learning pace can increase motivation. Adjacent to this, institutions must ensure comprehensive training for teachers and students so that AR tools are effective in classrooms (Huang et al., 2021). As Ly (2024) emphasizes the multifaceted roles of teachers in promoting learner-centered environments, it is imperative to equip educators with the skills needed to integrate AR technologies into their teaching strategies. This preparation will ensure that AR adoption aligns with pedagogical goals and enhances learning outcomes.

Research Implications

This review further establishes the benefits of AR beyond learning engagement and motivation in the process of acquiring language, signifying how important it is to design inclusive experiences for students. What can be drawn from the review is that this tool is predominantly tested with vocabulary learning, leaving a blind spot for its potential synergy with grammar and listening mastery (Economides et al., 2020). Our in-depth literature analysis has demonstrated this technology's efficacy, especially its interactive nature, in facilitating grammar and listening skills development through conversational practice with virtual entities (Chang et al., 2020; Huang et al., 2020). The integration of blended learning methodologies has demonstrated the potential to enhance flexibility and accessibility in English language teaching (Tran, 2024). However, it also highlights the need for comprehensive preparation, particularly regarding learners' readiness and technical competencies. These factors align with the challenges identified in the adoption of AR technologies, where successful implementation hinges on students' familiarity and comfort with technological tools. Future studies should continue with the exploration of more factors from AR that accumulate positive perceptions of students, teachers, and administrators, or in other words, elements that facilitate satisfaction, enjoyment, as well as measurable results. While engagement levels with AR tools have been acknowledged, quantitative studies are recommended to measure these effects more comprehensively.

Additionally, the recognition of AR's potential to foster intercultural competence (Liu et al., 2023b) warrants further research to answer the "how" question. AR-based instruction has indeed been shown to sharpen intercultural competence better than conventional methods. However, this aspect of AR's impact on cultural understanding and communication skills should be explored in greater depth. Furthermore, longitudinal studies to assess the long-term effects of AR on language learning outcomes through varied alternative instructional methods are also needed (Fombona et al., 2017).

Thus, this review enriches knowledge about AR's opportunities in ELT, the grounded approach to its inclusion for people with disabilities, AR's expansiveness toward various language skills, and the opportunities for its further development. In this case, when adopting the inclusive design principles and the use of appliances that may apply, one is able to use AR to enhance effective, accessible, and inclusive learning from the aspects of language learning with the overall learning needs of every learner.

Conclusion

This review highlights the transformative potential of Augmented Reality (AR) in English Language Teaching (ELT), particularly in promoting inclusivity and addressing diverse learning needs. The findings demonstrate that AR's multisensory approach effectively caters to visual, auditory, and kinesthetic learners, providing enhanced opportunities for engagement and personalized learning experiences. Through alignment with Universal Design for Learning (UDL) principles, AR fosters accessibility, cultural responsiveness, and equitable participation for learners of varying attributes and abilities. Additionally, AR's ability to reduce learning anxiety and improve motivation through interactive and gamified elements further underscores its value as a pedagogical tool.

Despite these benefits, challenges such as technological accessibility, teacher preparedness, and the long-term sustainability of AR's engagement effects remain top areas for further investigation. Notable gaps in the literature include the limited exploration of AR's impact on higher-order cognitive skills, its integration with other instructional technologies, and its potential to foster intercultural competence.

Previous studies show that the integration of inclusive design is vital for increasing the benefits associated with AR. Some of the main approaches that can be adopted include individualization, availability, permeability, and cultural sensitivity, which are necessary to adopt suitable and acceptable AR approaches for learners. However, when using such approaches, educators should adhere to the UDL principles to provide students with joyous, inclusive, and accessible experiences within AR contexts. Some advantages for learners derived from current AR studies and the existing research environment encompass the beneficial effects of those principles.

However, some disadvantages still exist that need to be taken into consideration. Therefore, more quantitative and follow-up studies are needed to establish the long-term impact of AR on other overall language learning achievements as well as other less explored skills. New investigations can establish ways of effectively promoting the use of AR in context with other approaches and materials that can make the learning process diverse and encompassing. Future studies should also investigate the broad impact of AR in relation to student groups that exhibit different learning styles. The attainment of such knowledge assists in the development of enhanced and particular AR interventions.

All in all, evidence demonstrates that AR has the enormous capacity to introduce alterations in the ELT context, derive the process from learner-centered analysis, embrace equality, and enhance learner outcomes. Moving beyond the limitations above and expanding the exploration of AR potentials will help researchers and educators pave the way towards a better future of foreign language learning.

References

Abascal, J., & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers, 17*(5), 484–505. https://doi.org/10.1016/j.intcom.2005.03.002

Aguayo, C., Cochrane, T., & Narayan, V. (2017). Key themes in mobile learning: Prospects for learner-generated learning through AR and VR. *Australasian Journal of Educational Technology, 33*(6), 27–40. https://doi.org/10.14742/ajet.3671

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review, 20*(1), 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Azuma, R. T. (1997). A survey of augmented reality. *Presence: teleoperators & virtual environments, 6*(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Integrating augmented reality in language learning: Pre-service teachers' digital competence and attitudes through the TPACK framework. *Education and Information Technologies, 27*(9), 12123-12146. https://doi.org/10.1007/s10639-022-11123-3

Beetham, H., & Sharpe, R. (Eds.). (2019). R*ethinking pedagogy for a digital age: Principles and practices of design.* Routledge.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.

Bistaman, I. N. M., Idrus, S. Z. S., & Abd Rashid, S. (2018, June). The use of augmented reality technology for primary school education in Perlis, Malaysia. In *Journal of Physics: Conference Series* (Vol. 1019, No. 1, p. 012064). IOP Publishing. https://doi.org/10.1088/1742-6596/1019/1/012064

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in education–cases, places and potentials. *Educational Media International, 51*(1), 1-15. https://doi.org/10.1080/09523987.2014.889400

Burgstahler, S. (2015). Opening doors or slamming them shut? *Online learning practices and students with disabilities. Social Inclusion, 3*(6), 69-79.

Cai, Y., Pan, Z., & Liu, M. (2022). Augmented reality technology in language learning: A meta-analysis. *Journal of Computer Assisted Learning, 38*(4), 929-945.

CAST. (2018). *Universal design for learning guidelines version 2.2*. Retrieved from http://udlguidelines.cast.org/static/udlg2.2-text-a11y.pdf

Chang, Y. S., Chen, C. N., & Liao, C. L. (2020). Enhancing English-learning performance through a simulation classroom for EFL students using augmented reality—A junior high school case study. *Applied Sciences, 10*(21), 7854. https://doi.org/10.3390/app10217854

Chen, W. (2020). Visual learning through augmented reality. *Journal of Educational Multimedia*, 19(4), 45-58. https://doi.org/10.3390/app10217854

Cheng, K. H., & Tsai, C. C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of science education and technology, 22*, 449-462. https://doi.org/10.1007/s10956-012-9405-9

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design

elements to gamefulness: defining" gamification". *In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15).

Do, T. T. L., Pham, N. T., & Ngo, P. A. (2024). Measuring EFL Learners' Perceptions of Technology Self-efficacy in Online Language Learning. *International Journal of Language Instruction*, *3*(3), 54–78. https://doi.org/10.54855/ijli.24334

Dunn, R., & Dunn, K. (1993). *Teaching secondary students through their individual learning styles: Practical approaches for grades 7-12*. Allyn and Bacon.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of science Education and Technology*, *18*, 7-22.

Economides, A. A., Parmaxi, A., & Demetriou, A. A. (2020). Augmented reality in language learning: A state-of-the-art review of 2014–2019. *Journal of Computer Assisted Learning*, *36*(6), 861–875. https://doi.org/10.1111/jcal.12486

Fombona, J., Pascual-Sevillana, A., & Gonzalez-Videgaray, M. (2017). M-learning and augmented reality: A review of the scientific literature on the WoS Repository. *Comunicar: Media Education Research Journal, 25*(2), 63-71.

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educational Research Review, 27*, 244-260.

Hall, T. E., Meyer, A., & Rose, D. H. (Eds.). (2012). *Universal design for learning in the classroom: Practical applications*. Guilford press.

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. *Tertiary assessment & higher education student outcomes: Policy, practice & research, 259*, 275.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Huang, X., Zou, D., Cheng, G., & Xie, H. (2021). A Systematic Review of AR and VR Enhanced Language Learning. *Sustainability, 13*(9), 4639. https://doi.org/10.3390/su13094639

Hsu, T. C. (2019). Effects of gender and different augmented reality learning systems on English vocabulary learning of elementary school students. *Universal Access in the Information Society, 18*, 315-325.

Ibáñez, M. B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education, 123*, 109-123.

Iqbal, M. Z., & Campbell, A. G. (2021, April). Investigating challenges and opportunities of the touchless hand interaction and machine learning agents to support kinesthetic learning in augmented reality. *In 26th International Conference on Intelligent User Interfaces-Companion* (pp. 99-101).

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior*, *25*(2), 332-338. https://doi.org/10.1016/j.chb.2008.12.019

Ke, F., & Hsu, Y. C. (2015). Mobile augmented-reality artifact creation as a component of mobile computer-supported collaborative learning. *The Internet and Higher Education*,

26, 33-41. https://doi.org/10.1016/j.iheduc.2015.04.003

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75-86.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. MIT Press. https://books.google.com.vn/books?hl=vi&lr=&id=I0kaFNaK704C&oi=fnd&pg=PR5#v=onepage&q&f=false

Klopfer, E., Squire, K., & Jenkins, H. (2002, August). Environmental detectives: PDAs as a window into a virtual simulated world. In *Proceedings. IEEE international workshop on wireless and mobile technologies in education* (pp. 95-98). IEEE. https://doi.org/10.1109/wmte.2002.1039227

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(3), 271-289. https://doi.org/10.1017/S0958344008000335

Ladau, E. (2021). *Demystifying disability: What to know, what to say, and how to be an ally*. Ten Speed Press.

Lee, J., & Park, H. (2020). Culturally responsive AR design in language learning. *Journal of Cultural Education*, *33*(2), 101-115.

Li, K. C., & Wong, B. T. M. (2021). A literature review of augmented reality, virtual reality, and mixed reality in language learning. *International Journal of Mobile Learning and Organisation*, *15*(2), 164-178. https://doi.org/10.1504/ijmlo.2021.114516

Liao, C. H. D., Wu, W. C. V., Gunawan, V., & Chang, T. C. (2024). Using an augmented-reality game-based application to Enhance Language Learning and Motivation of Elementary School EFL students: A comparative study in Rural and Urban Areas. *The Asia-Pacific Education Researcher, 33*(2), 307-319.

Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Journal of Educational Technology & Society, 18*(4), 486-497.

Lin, L., Wu, C., & Chen, X. (2020). Multimodal content representation in AR. *Educational Technology Research and Development*, 68(5), 234-251.

Liu, M., Horton, L., & Lee, J. (2023a). Gamification in augmented reality for language learning. *Journal of Educational Computing Research*, 59(1), 28-50.

Liu, S., Gao, S., & Ji, X. (2023b). Beyond borders: exploring the impact of augmented reality on intercultural competence and L2 learning motivation in EFL learners. *Frontiers in Psychology*, *14*, 1-15. https://doi.org/10.3389/fpsyg.2023.1234905

López Belmonte, J., Moreno-Guerrero, A. J., López Núñez, J. A., & Pozo Sánchez, S. (2019). Analysis of the productive, structural, and dynamic development of augmented reality in higher education research on the web of science. *Applied Sciences, 9*(24), 5306.

Ly, C. K. (2024). Teachers' roles on English language teaching for promoting learner-centered language learning: A theoretical review. *International Journal of TESOL & Education, 4*(2), 78-98. https://doi.org/10.54855/ijte.24425

Manna, M. (2023). Teachers as augmented reality designers: a study on Italian as a foreign language–teacher perceptions. *International Journal of Mobile and Blended Learning*

(IJMBL), *15*(2), 1-16. https://doi.org/10.4018/ijmbl.318667

Marrahí-Gómez, V., & Belda-Medina, J. (2022). The application of augmented reality (AR) to language learning and its impact on student motivation. *International Journal of Linguistics Studies*, *2*(2), 07-14. https://doi.org/10.32996/ijls.2022.2.2.2

Martín-Gutiérrez, J., Fabiani, P., Benesova, W., Meneses, M. D., & Mora, C. E. (2015). Augmented reality to promote collaborative and autonomous learning in higher education. *Computers in human behavior, 51*, 752-761.

Meyer, A., Rose, D. H., & Gordon, D. (2014). Universal design for learning: Theory and practice, Wakefield MA: CAST.

Min, W., & Yu, Z. (2023). A bibliometric analysis of augmented reality in Language Learning. *Sustainability*, *15*(9), 7235. https://doi.org/10.3390/su15097235

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321-1329. Search.ieice.org. https://search.ieice.org/bin/summary.php?id=e77-d_12_1321

Nicholson, S. (2015). A recipe for meaningful gamification. *Gamification in education and business/Springer.*

Pasfield-Neofitou, S. (2014). Language learning and socialization opportunities in game worlds: Trends in first and second language research. *Language and Linguistics Compass*, *8*(7), 271-284. https://doi.org/10.1111/lnc3.12083

Pachler, N., Cook, J., & Bachmair, B. (2010). Appropriation of mobile cultural resources for learning. *International Journal of Mobile and Blended Learning (IJMBL)*, *2*(1), 1-21.

Pellas, N., Fotaris, P., Kazanidis, I., & Wells, D. (2019). Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning. *Virtual Reality*, *23*(4), 329-346. https://doi.org/10.1007/s10055-018-0347-2

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version, 1*(1), b92.

Puentedura, R. R. (2013). SAMR: Getting to transformation. *Retrieved May, 31*, 265-283.

Qiu, X. Y., Chiu, C. K., Zhao, L. L., Sun, C. F., & Chen, S. J. (2023). Trends in VR/AR technology-supporting language learning from 2008 to 2019: A research perspective. *Interactive Learning Environments*, *31*(4), 2090-2113. https://doi.org/10.1080/10494820.2021.1874999

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Association for Supervision and Curriculum Development, 1703 N. https://eric.ed.gov/?id=ED466086

Sannikov, S., Zhdanov, F., Chebotarev, P., & Rabinovich, P. (2015). Interactive educational content based on augmented reality and 3D visualization. *Procedia Computer Science*, *66*, 720-729. https://doi.org/10.1016/j.procs.2015.11.082

Seale, J. (2013). *E-learning and disability in higher education: accessibility research and practice*. Routledge.

Şimşek, B., & Direkçi, B. (2023). The effects of augmented reality storybooks on student's

reading comprehension. *British Journal of Educational Technology*, *54*(3), 754 772. https://doi.org/10.1111/bjet.13293

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. ASCD. https://books.google.com.vn/books?hl=vi&lr=&id=CLigAwAAQBAJ&oi=fnd&pg=PP1&dq=Tomlinson

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. https://doi.org/10.54855/ijte.24426

UNESCO. (2005). *Guidelines for inclusion: Ensuring access to education for all*. UNESCO.

Wedyan, M., Falah, J., Elshaweesh, O., Alfalah, S. F. M., & Alazab, M. (2022). Augmented Reality-Based English Language Learning: *Importance and State of the Art. Electronics, 11*(17), 2692. https://doi.org/10.3390/electronics11172692

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & education*, *62*, 41-49. https://doi.org/10.1016/j.compedu.2012.10.024

Wu, P. (2019). Audio integration in AR for language learning. *Journal of Language Teaching and Research*, 10(3), 477-486.

Yen, J. C., Tsai, C. H., & Wu, M. (2013). Augmented reality in the higher education: Students' science concept learning and academic achievement in astronomy. *Procedia-social and behavioral sciences*, *103*, 165-173. https://doi.org/10.1016/j.sbspro.2013.10.322

Yulian, R., Ruhama, U., & Sucipto, S. (2022). Developing Augmented Reality (AR) as Assisted Technology in Reading Based on Content-Language Integrated Learning. *Jurnal Teknologi Pendidikan*, *24*(1), 23-37.

Zhao, G., Zhang, Q., Chu, J., Li, Y., Liu, S., & Lin, L. (2018, November). Augmented reality application for plant learning. *In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 1108-1111). IEEE.

Biodata

Ms. Ngo Nguyen Thien Duyen completed her MA in Applied Linguistics and TESOL at Macquarie University, Australia. She has published articles in local and international journals. She is currently a lecturer at Ho Chi Minh City University of Economics and Finance. Her research interests include but are not limited to TESOL, teacher professional development, assessment practice, and intercultural communication.

Mr. Vo Trong Nghia completed his Master Degree in TESOL at Monash University, Australia. He is currently a lecturer at Ho Chi Minh City University of Economics and Finance. His interests include but are not limited to effective English classroom practices and integration of technology in teaching and learning.

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.

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