



USING CHATGPT TO DESIGN ENGLISH PRACTICE TESTS FOR THE NATIONAL HIGH SCHOOL GRADUATION EXAM IN VIETNAM

Nguyen Thi Hong Minhⁱ,
Dinh Dam Hai Anh,
Vu Hoang Dung

Faculty of Foreign Languages Education,
Thai Nguyen University of Education,
Vietnam

Abstract:

The 2025 National High School Graduation Exam in Vietnam introduces changes to the English test format with more diverse question types, which may lead to some difficulties for both teachers and students regarding teaching and learning documents for practice. This study explores the use of ChatGPT in generating English practice tests aligned with the new 2025 national high school graduation exam format and evaluates the quality of the generated exams in terms of reliability, validity, and difficulty. Conducted at two high schools in a northern province of Vietnam, the research involved 85 to 154 students in grade 12 and followed five stages: designing prompts, gathering expert feedback, revising prompts, implementing the test with students, and evaluating outcomes. The practice test design process was carried out through eight stages: creating a test matrix, developing prompt templates, testing and refining prompts, storing as a validated template, including topics from the 2018 syllabus, creating questions for each part, reviewing and refining each part, and compiling and formatting the final test. A combination of expert judgment and statistical analysis was employed to evaluate the quality of the test. Expert feedback assessed face validity and content validity through Likert questionnaires. Students' practice test results were analyzed using SPSS to examine construct validity, reliability (Cronbach's alpha), and difficulty of items, tests. Findings show that with refined prompts, ChatGPT can create reasonably valid, reliable, and difficult practice tests by using the 8 stages of practice test design.

Keywords: ChatGPT, English practice tests, National High School Graduation Exam, AI in education, prompt engineering, test design

ⁱ Correspondence: email minhnhth@tnue.edu.vn

1. Introduction

English has become an essential subject in Vietnam's general education system, particularly in the context of globalization and national educational reform (Hien & Loan, 2018; Van Van, 2010). Recognizing its importance, the Ministry of Education and Training (MOET) has implemented competency-based curriculum reforms and is revising the National High School Graduation Exam to better align with real-life communication and critical thinking skills. However, the implementation of these reforms presents significant challenges. Teachers must design materials and practice tests that meet new requirements, often with limited time and professional development (Le & Do, 2019). Moreover, the persistent urban-rural gap in teaching quality and access to digital tools continues to hinder educational equity (Chen, 2015; Mathew, 2015).

Recent advancements in artificial intelligence offer potential solutions to these challenges. Large language models, such as OpenAI's ChatGPT, have demonstrated the ability to generate domain-specific content, including exam-style tasks, and to support autonomous learning through immediate, contextual feedback (Opara, Theresa, & Aduke, 2023). These capabilities suggest promising applications in English language assessment, particularly in helping teachers create practice tests that reflect updated exam formats. However, caution is warranted. Concerns regarding the accuracy, pedagogical relevance, and ethical use of AI-generated content must be addressed through structured guidance and teacher oversight (Nguyen & Tran, 2024; Le, 2024).

This study aims to develop a clear and practical process for using ChatGPT to generate English practice tests aligned with the revised format of the National High School Graduation Exam in Vietnam. It also evaluates the quality of these tests in terms of reliability, validity, and difficulty. Specifically, the study seeks to answer two research questions:

- 1) What steps and prompts can be used to guide ChatGPT in creating appropriate English practice tests for the Vietnamese context?
- 2) How do the AI-generated tests perform regarding their psychometric properties and educational usefulness?

2. Methods of Study

The study engaged a varied group of high school students and experienced English language educators from northern Vietnam to examine how effective the English practice test was. Conducted over a ten-week period, the study involved between 85 and 154 students from three classes at two high schools, all of whom had voluntarily selected English as a core subject for their national examination. To enrich the dataset with professional perspectives, two distinct groups of educators were included. During the Trial Practice Test (TPT) phase, 30 high school English teachers participated. The Final Practice Test (FPT) phase involved 40 participants, encompassing both high school teachers and university lecturers, most possessing over two decades of teaching

experience and substantial familiarity with the national exam framework. Notably, many of these educators integrated AI tools into their instructional practices and assessment design processes.

The study followed a five-stage design process:

- 1) developing prompts and test blueprints aligned with national standards;
- 2) conducting an expert survey to evaluate a trial test;
- 3) finalizing prompt design and producing the first full test;
- 4) implementing weekly tests over ten weeks in a real classroom setting; and
- 5) conducting final expert evaluations and statistical analysis of test data.

This study adopts a mixed-method approach, combining quantitative and qualitative techniques to evaluate the reliability, validity, and difficulty of English tests generated by ChatGPT. Quantitative data, including student test scores across ten practice tests, were analyzed using SPSS to compute indicators such as Cronbach's alpha, Corrected Item-Total Correlation (CITC), and Item/Test Difficulty Index (IDI/TDI). Qualitative insights were collected through expert questionnaires employing both Likert-scale and open-ended items, allowing for detailed judgments on face and content validity.

Validity was assessed through both construct and content validity. Items with CITC values below 0.30 were considered weak and subject to revision or removal (Hassan, 2023). Content validity was evaluated using CVI: values ≥ 0.80 indicated acceptable items; scores between 0.70 and 0.79 suggested revision; and scores < 0.70 required elimination (Hennus *et al.*, 2021). Face validity was interpreted using Likert-scale mean scores, following Sözen's (2019) classification.

Reliability was measured by Cronbach's Alpha, with values ≥ 0.70 considered acceptable, reflecting internal consistency of test items and expert evaluations (Tavakol & Dennick, 2011).

Kline (2015) notes that the mean p-value index ranges from 0 (very difficult) to 1 (very easy), and Brown (2022) considers approximately 0.5 as the ideal level for balancing test difficulty.

3. Findings

3.1 The Procedures for Designing English Tests Using ChatGPT

A process of 8 stages for designing English tests with the use of ChatGPT was developed and presented in the following diagram.

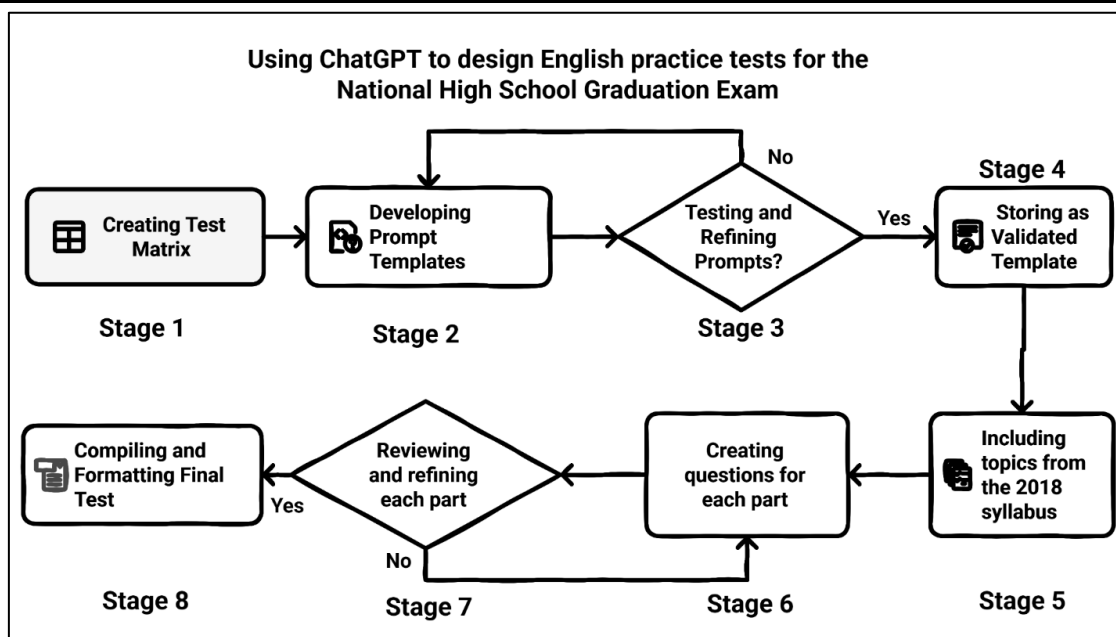


Figure 1: The 8-stage Process of Designing English Tests Using ChatGPT

Stage 1: Creating test matrix

In the first stage of the process, official exam samples were analyzed to determine test structure, question types, difficulty levels, and frequent topics. The gained information was utilized to build a foundational test matrix for prompt design.

Stage 2: Developing Prompt Templates

A standardized prompt format was designed with five elements: Type, Level, Instruction, Topic/Theme, and Example.

Stage 3: Testing and Refining Prompts

In the third stage of the process, each prompt was tested with ChatGPT at least 10 times to evaluate correctness, clarity, difficulty, and formatting. Revisions were made as needed.

Stage 4: Storing as Validated Template

It is advised that well-performing prompts be archived in categorized collections by test section for long-term reuse in practice test creation.

Stage 5: Including Topics from the 2018 Syllabus

The 2018 English curriculum was reviewed to identify key themes, which were then integrated into the prompts to ensure content alignment.

Stage 6: Creating Questions for Each Part

It is expected that each validated prompt be used to generate questions for individual exam sections, maintaining logical consistency and alignment with test requirements.

Stage 7: Reviewing and Refining Each Part

The generated sections were carefully examined, and errors in content, structure, or difficulty level were corrected to ensure high-quality test output.

Stage 8: Compiling and Formatting the Final Test

In the final stage of the process, the complete test was compiled and formatted according to MOET standards (e.g., bold answers, aligned text, clear instructions) to simulate a real exam experience.

3.2 Test Quality

This study evaluates the overall quality of English practice tests generated by ChatGPT based on three fundamental measurement criteria: validity, reliability, and difficulty. Specifically, validity is examined across three dimensions: construct, content, and face validity. Reliability is measured using Cronbach's Alpha, and difficulty is assessed through both the item difficulty index and the test difficulty index. These indicators provide a comprehensive and evidence-based evaluation of the tests' effectiveness and educational value.

3.3 Reliability

The table below presents the reliability coefficients (Cronbach's Alpha) of the ten English practice tests generated by ChatGPT as part of the test development process.

Table 1: Cronbach's Alpha of ChatGPT-generated English Practice Tests

Cronbach's Alpha					
Practice Test 1	Practice Test 2	Practice Test 3	Practice Test 4	Practice Test 5	N of Items
.816	.764	.728	.753	.752	40
Practice Test 6	Practice Test 7	Practice Test 8	Practice Test 9	Practice Test 10	N of Items
.865	.814	.830	.833	.863	40

Cronbach's Alpha coefficients were calculated for ten English practice tests generated by ChatGPT. All tests met the acceptable reliability threshold ($\alpha \geq 0.70$), with an average of 0.792. The first five tests, developed before refinement, showed lower reliability ($\alpha = 0.728$ – 0.816), while the last five, created after targeted revisions, demonstrated significantly higher consistency ($\alpha = 0.814$ – 0.865). These results confirm that refinement substantially improves the reliability of AI-generated test items.

3.4 Validity

3.4.1 Construct Validity

The table below presents the results of expert validation regarding the content validity of the English practice tests generated by ChatGPT.

Table 2: Number of Items Categorized by CITC Value Range

CITC Value Range	PT 1	PT 2	PT 3	PT 4	PT 5
< 0.30	21	29	34	29	29
0.30 – 0.50	18	6	6	11	9
> 0.50	1	5	0	0	2
CITC Value Range	PT 6	PT 7	PT 8	PT 9	PT 10
< 0.30	22	24	22	20	19
0.30 – 0.50	5	5	12	19	8
> 0.50	13	11	6	1	13

Across all ten ChatGPT-generated practice tests, item quality, measured by CITC, varied considerably. Tests 6, 7, and 10 had the highest number of strong items (CITC > 0.50), each with 11 to 13 items, reflecting high internal consistency and test coherence. In contrast, Tests 3, 4, and 8 had no or very few strong items, with weak items (CITC < 0.30) dominating over half the test content. This indicates potential reliability concerns and highlights the need for item-level revision.

Moderately correlated items (CITC 0.30–0.50) were fairly common across tests, particularly in PT 9, which had 19 such items. These items suggest acceptable, but improvable, quality. Notably, PT 3 had the poorest profile overall, with 34 weak items and none in the strong range, while PT 10 combined strong internal consistency with a relatively low number of weak items (19), showcasing the effectiveness of refinement. Overall, the distribution reveals a trend: earlier tests (PT 1–5) generally had more weak and fewer strong items, while later, revised tests (PT 6–10) showed improvement in item strength. This reinforces the value of prompt refinement and expert review in enhancing AI-generated assessments.

3.4.2 Content Valid

The table below compares the content validity index (S-CVI/Ave) scores of the Trial and Final Practice Tests across seven parts. It highlights a significant improvement in item relevance from the trial to the final version of the test.

Table 3: Average Content Validity Index (S-CVI/Ave) between the TPT and FPT

Trial Practice Test	S-CVI/Ave	Final Practice Test	S-CVI/Ave
Part I	0.77	Part I	0.86
Part II	0.79	Part II	0.85
Part III	0.81	Part III	0.90
Part IV	0.81	Part IV	0.89
Part V	0.79	Part V	0.85
Part VI	0.81	Part VI	0.89
Part VII	0.76	Part VII	0.91
Overall S-CVI/Ave	0.79	Overall S-CVI/Ave	0.88

The Trial Practice Test (TPT) was evaluated for its content validity by an expert panel, with 36 out of 70 items meeting or exceeding the 0.80 threshold for relevance. This means that around 51.4% of the items were judged to be valid. However, 34 items (48.6%) scored

below the threshold, indicating that some sections, especially Part VI, required revisions to meet content standards.

Part III, IV, and VI scored the highest (0.81), suggesting they were nearly valid, with minimal revisions needed. In contrast, Parts I, II, and V scored closer to 0.79, needing minor adjustments. Part VII, scoring the lowest at 0.76, required the most significant changes due to its complex structure.

The Final Practice Test (FPT) reveals a remarkable enhancement in content quality and expert consensus compared to the earlier Trial version. Out of the 70 total I-CVI scores assessed across 10 questions and 7 parts, all 70 individual CVI scores in the FPT phase reached or exceeded the 0.80 threshold, indicating a 100% validation rate for item relevance. This is a clear indication that the content no longer requires revision and confirms the success of the test development and refinement process. In contrast to the TPT, where only 36 of the 70 items (51.4%) met the validity threshold and 34 required revision, the FPT demonstrates a test structure that is both educationally coherent and statistically sound.

3.4.3 Face Validity

The table below presents the average expert ratings for each part of the Trial and Final Practice Tests. It shows a consistent improvement in the perceived quality of the test after revisions, as reflected in higher mean scores across all sections.

Table 4: Expert Ratings of Trial and Final Practice Tests Based on Likert Scale Evaluations

Trial Practice Test	Mean/Ave		Final Practice Test	Mean/Ave	
Part I	3.92	Agree	Part I	4.08	Agree
Part II	3.93	Agree	Part II	4.04	Agree
Part III	3.93	Agree	Part III	4.12	Agree
Part IV	3.98	Agree	Part IV	4.08	Agree
Part V	3.87	Agree	Part V	4.07	Agree
Part VI	3.88	Agree	Part VI	4.10	Agree
Part VII	3.85	Agree	Part VII	4.14	Agree
Overall Mean/Ave	3.91	Agree	Overall Mean/Ave	4.09	Agree

All expert ratings for both the trial and final practice tests fall within the “Agree” range, indicating a consistent consensus among experts on the face validity of the test across all sections.

A direct comparison between the TPT and FPT reveals that the final version of the test was rated more positively across multiple indicators. One key metric is the number of items achieving a mean score of 4.0 or higher, representing strong agreement from experts. In the TPT, 28 out of 91 items reached this threshold, while in the FPT, 69 out of 91 items achieved a mean score of at least 4.0. This represents an increase of 41 items, suggesting that the revisions incorporated into the FPT effectively enhanced the perceived validity of the test.

When comparing the mean scores of each part between the TPT and FPT, improvements were observed consistently. For Part I, the mean increased from 3.92 in the TPT to 4.08 in the FPT. Similarly, Part II improved from 3.93 to 4.04, and Part III from 3.93 to 4.12. The largest increase was observed in Part VII, which rose from 3.85 to 4.14, marking a difference of +0.29 points. Even the parts that were already relatively well-rated in the TPT, such as Part IV (from 3.98 to 4.08), showed gains. These improvements indicate that revisions were effective across the tests, but especially significant in Part VII.

The overall mean scores also highlight this positive trend. The TPT achieved an overall mean of 3.91, while the FPT reached 4.09, an increase of 0.18 points. This suggests a general enhancement in the test's perceived face validity following expert feedback and subsequent modifications.

In addition to face validity, the reliability of the questionnaire was assessed using Cronbach's Alpha.

Table 5: Cronbach's Alpha of the Questionnaire

Cronbach's Alpha	TPT	FPT
	.933	.993
N of Items	91	

The results show high reliability coefficients: 0.933 for the TPT and 0.993 for the FPT. Both values exceed the commonly accepted threshold of 0.9, indicating excellent internal consistency among the questionnaire items. The increase in Cronbach's Alpha from the TPT to the FPT suggests that the final version of the test not only improved in perceived validity but also in the consistency of expert evaluations. This high reliability reinforces the credibility of the expert feedback and supports the validity of the findings reported in this study.

3.5 Difficulty

3.5.1 Test Difficulty Index (TDI)

Based on the data drawn from the Test Difficulty Index (TDI), the overall difficulty level of the ten English practice tests designed using ChatGPT is appropriate and pedagogically sound.

The ten practice tests analyzed exhibit ideal difficulty levels, with Test Difficulty Indices (TDIs) falling within the optimal range of 0.30 to 0.70, which is considered moderate and effective for educational assessments. None of the tests were either too easy or too challenging, ensuring balanced difficulty levels that promote student engagement while maintaining test validity.

Table 6: Test Difficulty Index (TDI) of ChatGPT-generated English Practice Tests

	N	Minimum	Maximum	Mean	TDI	Std. Deviation
Practice Test 1	154	5	40	18.40	0.46	6.224
Practice Test 2	142	10	32	22.22	0.56	5.061
Practice Test 3	134	9	29	21.75	0.54	3.950
Practice Test 4	113	11	36	26.58	0.66	4.336
Practice Test 5	107	6	31	23.91	0.60	4.743
Practice Test 6	96	10	33	25.48	0.64	4.540
Practice Test 7	93	5	33	25.33	0.63	4.756
Practice Test 8	93	10	34	26.51	0,66	4.536
Practice Test 9	90	5	31	24.62	0,62	4.024
Practice Test 10	88	7	30	22.76	0,57	4.389

Test 1, with the lowest TDI of 0.46, is the most difficult, showing a wider variation in scores, possibly due to its challenging items. It may be best used for diagnostic purposes or with advanced learners. On the other hand, Tests 4 and 8, with TDIs of 0.66, are the easiest, offering higher mean scores and more consistent performance, making them suitable for formative assessment or building confidence in students.

The remaining tests (2, 3, 5, 6, 7, 9, and 10) show a more consistent range of TDIs between 0.54 and 0.64, providing a balanced progression of difficulty. Test 3 stands out for having the lowest standard deviation, suggesting uniform performance, while Tests 5, 6, 7, and 9 offer moderate challenges with slight variations in difficulty.

3.5.2 Item Difficulty Index

The table below illustrates the distribution of cognitive levels across the MOET benchmark and ten practice tests, focusing on Recognition, Understanding, and Application. It highlights a clear progression in test quality, with later practice tests aligning more closely with MOET standards.

Table 7: Cognitive level distribution across MOET and practice tests

Scale	MOET	(%)	PT 1	PT 2	PT 3	PT 4	PT 5
Recognition	6	15	2	8	6	12	15
Understanding	18	45%	25	27	27	26	21
Application	16	40%	13	5	7	2	4
Total	40	100%	40	40	40	40	40
	MOET	(%)	PT 6	PT 7	PT 8	PT 9	PT 10
Recognition	6	15%	10	11	14	13	12
Understanding	18	45%	13	14	12	13	12
Application	16	40%	17	15	14	14	16
Total	40	100%	40	40	40	40	40

The table compares the cognitive-level distribution of questions—Recognition, Understanding, and Application—in the MOET test and ten practice tests (each with 40 questions). The MOET test serves as a benchmark, with 15% Recognition, 45% Understanding, and 40% Application, emphasizing higher-order thinking.

Practice Tests 1–5 show major imbalance: Test 1 underrepresents Recognition and overuses Understanding; Test 2 overuses both Recognition and Understanding but lacks Application; Test 3 matches MOET in Recognition but underperforms in Application; Test 4 and Test 5 overemphasize Recognition and fail to include sufficient Application items, resulting in low cognitive challenge.

From Practice Test 6 onward, the design improves. Test 6 exceeds the MOET standard in Application (43%) and shows a more even distribution. Test 7 continues this trend with strong Application content. Test 8 maintains acceptable Application but overuses Recognition. Test 9 achieves near balance across all levels. Test 10 shows the best alignment with MOET, matching it in Application (40%) and offering a mature, well-structured assessment.

This progression highlights improved test design over time. It also notes that perceived difficulty depends not only on test structure but also on student mindset. Early-stage practice may yield poor performance due to low engagement, while later-stage tests provide more accurate reflections as students become more serious.

4. Discussion

Using generative AI tools like ChatGPT in education, especially for creating practice tests, is still a new area. While many studies show ChatGPT can support English learning, its outputs require careful review. In this study, a prompt-based framework was used to generate ten complete English practice tests mirroring the Vietnamese National High School Graduation Exam. This supports Xiao *et al.* (2023), who found that with well-designed prompts, language models can create coherent tasks. Mollick (2023) also highlighted the effectiveness of prompt engineering in aligning AI output with learning goals, a view confirmed by teacher feedback.

Expert evaluations showed the tests had high face and content validity, with appropriate format, language, and relevance, similar to findings by Hongxia and Razali (2025), who noted the structural quality and learner engagement in ChatGPT outputs. Student performance further confirmed the tests' reliability and suitable difficulty across ability levels.

Unlike earlier studies focusing on specific skills, this research applied a full-test approach using a five-part prompt formula (Type – Level – Instruction – Topic/Theme – Example). This ensured consistency and allowed teachers to generate quality materials without advanced tools.

Educators noted that teacher review remains essential for refining AI-generated items, echoing Mollick's (2023) view that AI should support, not replace, pedagogical expertise. Most participants found ChatGPT helpful for drafting multiple-choice questions, consistent with Nguyen (2023), though this study went further by analyzing test validity through actual performance data and expert feedback.

Nonetheless, early outputs exhibited limitations, such as incomplete distractors, topic drift, and inconsistent formatting, echoing issues documented by McGee (2024) in

a study on Microsoft Copilot. Similar problems were observed here, but through prompt refinement and formatting guidelines, the quality of items improved significantly, reducing manual editing. Even so, final item validation still depended on expert teacher review, reaffirming that human judgment remains essential in AI-assisted assessment design.

5. Conclusions

This study examined the use of ChatGPT in generating English practice tests for the National High School Graduation Exam in Vietnam. The results show that the effectiveness of ChatGPT depends largely on how prompts are designed and revised. With clear, well-structured prompts and careful human adjustments, ChatGPT can produce test items that are relevant and aligned with exam requirements.

The generated tests were found to meet essential criteria: reliability, validity, and appropriate difficulty. Reliability was at an acceptable level, showing consistency in assessing student performance. Content and face validity were generally strong, with expert reviews confirming alignment with the Ministry of Education's standards. However, construct validity may be influenced by both subjective and objective factors, such as students' mindset and effort during different preparation phases, making it challenging to assess true test difficulty based solely on item design. In terms of difficulty, both test difficulty index and item difficulty index were appropriate, matching the expectations set by the Ministry. This balance ensures the tests are suitable for high school students preparing for the national exam.

In conclusion, ChatGPT can serve as a helpful tool for teachers in test design if used with clear goals and expert oversight. Further research could explore ways to enhance construct validity and apply AI in more stages of assessment development.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Author(s)

Dr. Nguyen Thi Hong Minh is currently a lecturer of English language education at Thai Nguyen University of Education. Her research interests include innovation practices in foreign language teaching and learning, professional development for in-service teachers and English curriculum design and evaluation. Recently she has been motivated by the integration of technologies in English instruction, including the use of AI for learners' proficiency development. She was recognized by Ministry of Education and Training for her supervision of students' researcher papers which were awarded 3rd prize in 2023 and 2024.

Dinh Dam Hai Anh is an undergraduate student majoring in English language education at the Faculty of Foreign Languages Education, Thai Nguyen University of Education. With a strong interest in exploring and applying artificial intelligence in English language

learning and teaching, he has actively volunteered in various academic projects, including the TEMA project (Teaching English Multilingually through Art).

Vu Hoang Dung is currently studying English language education at the Faculty of Foreign Languages Education, Thai Nguyen University of Education. With an interest in the technology and pedagogy, she is especially keen on applying artificial intelligence to enhance English learning experiences. Dung has also taken part in academic and extracurricular projects as an active volunteer, most notably her involvement in the TEMA (Teaching English Multilingually through Art).

References

- Brown, T. (2022). Analysis of mathematics test difficulty in secondary education. *Journal of Educational Measurement*, 59(1), 45-58. <https://doi.org/10.1002/jem.1234>
- Chen, C. (2015). Planning and implementation of elementary school English education in Taiwan. In B. Spolsky & K. Sung (Eds.), *Primary school English-language education in Asia: From policy to practice* (pp. 89–104). Routledge. Retrieved from <https://www.routledge.com/Primary-School-English-Language-Education-in-Asia-From-Policy-to-Practice/Spolsky-Moon/p/book/9780415629683?srsId=AfmBOopQBPOxc01RQU52-NXhDeDpfUPhvv0Rf9r3Q23y-uQyNnCb4e5>
- Hassan, M. (2023). Construct validity analysis with Messick validity approach and Rasch model application on scientific reasoning test items. *Educational and Psychological Measurement*, 83(5), 789–812. <https://doi.org/10.1177/00131644221123456>
- Hien, T. T. T., & Loan, M. T. (2018). Current challenges in the teaching of tertiary English in Vietnam. In J. Albright (Ed.), *English tertiary education in Vietnam* (pp. 40–53). Routledge. <https://doi.org/10.4324/9781315212098>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press. Retrieved from <https://psycnet.apa.org/record/2015-56948-000>
- Le, V. C. (2024). An error analysis of the writing of 7th graders in Hanoi. *Vietnamese Journal of Language Studies*, 12(1), 45-62. <https://doi.org/10.17507/tpls.1411.27>
- Le, V. C., & Do, T. M. C. (2019). How current policies affect Vietnamese EFL teacher educators' professional development. *Language Policy*, 18(3), 387–405.
- Mathew, R. (2015). Young learner English language policy and implementation: A view from India. In B. Spolsky & K. Sung (Eds.), *Primary school English-language education in Asia: From policy to practice* (pp. 105–120). Routledge. Retrieved from <https://www.routledge.com/Primary-School-English-Language-Education-in-Asia-From-Policy-to-Practice/Spolsky-Moon/p/book/9780415629683?srsId=AfmBOopQBPOxc01RQU52-NXhDeDpfUPhvv0Rf9r3Q23y-uQyNnCb4e5>
- Nguyen, T. H., & Tran, L. M. (2024). Student perceptions of interaction in English-medium instruction classes at a Vietnamese technical university. *Journal of Asian*

- Education*, 15(2), 112-130. Retrieved from <https://doi.org/10.63023/2525-2445/jfs.ulis.5251>
- Opara, E., Theresa, A., & Aduke, T. C. (2023). ChatGPT for teaching, learning and research: Prospects and challenges. *Global Academic Journal of Humanities and Social Sciences*, 5(2), 33–47. Retrieved from <http://dx.doi.org/10.36348/gajhss.2023.v05i02.001>
- Van Van, H. (2010). The current situation and issues of the teaching of English in Vietnam. *Ritsumeikan Studies in Language and Culture*, 22(1), 7–18. Retrieved from https://www.researchgate.net/publication/326718789_The_Current_Situation_and_Issues_of_the_Teaching_of_English_in_Vietnam

Creative Commons licensing terms

Authors will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Applied Linguistics Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflict of interests, copyright violations and inappropriate or inaccurate use of any kind content related or integrated on the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).