

Perceptions of CMRU-TEP: English Majors versus Non-English Majors' Voices

PANACHANOK CHANWAIWIT

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University, Thailand

LALIDA WIBOONWACHARA*

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University, Thailand

Corresponding author email: lalida_wib@cmru.ac.th

| Article information | Abstract |
|---------------------------------|---|
| Article history: | <i>Chiang Mai Rajabhat University Test of English Proficiency (CMRU-TEP)</i> |
| Received: 19 Dec 2023 | <i>is a required English proficiency test for all CMRU students before</i> |
| Accepted: 8 Apr 2025 | <i>graduation. Despite its meticulous design, there is an opportunity for</i> |
| Available online: 17 Apr 2025 | <i>students to improve their scores through focused efforts and targeted</i> |
| | <i>support. This study employs an explanatory sequential mixed-methods</i> |
| Keywords: | <i>design, utilizing surveys and interviews, to explore student perceptions</i> |
| <i>Test perception</i> | <i>of CMRU-TEP and to propose improvement strategies to enhance test</i> |
| <i>Test usefulness</i> | <i>performance. Guided by Bachman and Palmer's (1996) model, the study</i> |
| <i>CMRU English test</i> | <i>involves 1,037 fourth-year students, including 155 English majors and</i> |
| <i>English proficiency test</i> | <i>882 non-English majors. Both groups consider CMRU-TEP "useful",</i> |
| | <i>addressing six qualities of test usefulness. The perceptions of majors and</i> |
| | <i>non-majors are similar. To enhance CMRU-TEP, the following recommendations</i> |
| | <i>are proposed: 1) Develop a test administration handbook. 2) Integrate</i> |
| | <i>a writing and speaking portfolio as part of the proficiency assessment.</i> |
| | <i>3) Ensure the test aligns with the focus on communicative skills. 4) Shorten</i> |
| | <i>test time and tasks. 5) Design tasks to stimulate real-world language</i> |
| | <i>use. 6) Explore the potential of computer-based testing as an alternative.</i> |
| | <i>The paper concludes by proposing tailored support for English and</i> |
| | <i>non-English majors based on identified needs.</i> |

INTRODUCTION

In Thailand, the government has emphasized the importance of English proficiency, along with other skills like ICT and global awareness, to foster a competitive workforce and achieve high-income country status, aligning with national plans such as Thailand 4.0 and the Thirteenth National Economic and Social Development Plan (2023-2027) (Office of the National Economic and Social Development Board, 2022; Wittayasin, 2017). To meet the goals set by national plans and prepare Thai students for the global workforce, the Common European Framework of Reference for Languages (CEFR) has been introduced as the framework guiding English language education in various aspects like curriculum, teaching methods, assessments, and teacher evaluations (Waluyo, 2019).

The Commission on Higher Education Standards (2024) has issued and disseminated the English language proficiency enhancement national policy to all tertiary institutions in Thailand. To provide flexibility to the institutions, individual higher education institutions have the authority to conduct their own in-house English language proficiency tests based on the CEFR or other standardized tests. University students take the language proficiency test as a requisite requirement for graduation to determine their individual levels of English language proficiency.

Thirty-eight Rajabhat universities nationwide, including Chiang Mai Rajabhat University (CMRU), have adopted the national policy (Charttrakul & Damnet, 2021). CMRU has implemented internal educational management policies focusing on fourth- and fifth-year students, aiming for a B1 proficiency level for all students. As part of this initiative, students are required to take the Chiang Mai Rajabhat University Test of English Proficiency (CMRU-TEP) (Chiang Mai Rajabhat University, 2023). However, the CMRU-TEP results from 2019 to 2022 show that less than 16 percent of these students have reached the B1 level, according to the CMRU-TEP score report for final-year undergraduate students (Chiang Mai Rajabhat University Language Center, 2023).

The persistent shortfall in meeting the university's language proficiency benchmarks underscores the need to refine CMRU-TEP. This involves investigating the test's flaws to enable students to improve their scores through more focused efforts and targeted support. Recently, CMRU-TEP has undergone revision, highlighting the importance of involving stakeholders, particularly the test-takers, in the preliminary process of test development. As suggested by Jackson (2022) and Zhou and Yoshitomi (2019), involving the perceptions of test-takers is an essential aspect of test development.

The present study aims to investigate the perceptions held by CMRU-TEP test-takers regarding the usefulness of CMRU-TEP and to explore beneficial and valuable insights that can be leveraged to enhance both the test itself and other related pedagogical practices and materials, aiding students in achieving the established standards set forth by the university.

Research objectives

1. To investigate usefulness of Chiang Mai Rajabhat University Test of English Proficiency (CMRU-TEP) based on the perceptions of English majors and non-English majors.
2. To propose strategies for enhancing CMRU-TEP and providing support for improving the test performance of CMRU students.

LITERATURE REVIEW

Perceptions

Perception refers to the process of organizing, identifying, and interpreting sensory input to create mental representations and meaningful experiences (Kasschau, 2003; Rohman et al.,

2019; Schacter et al., 2009). According to van de Watering et al. (2008), test-takers' perceptions refer to how students perceive the utility of tests or assessments they encounter in their courses to assist in achieving their goals. This aspect is typically studied in the realm of educational psychology, e.g., examining how students deal with anxiety, learning difficulties, or manage their learning processes. As noted by Hu (2022), test-takers may adapt their learning strategies based on these perceptions when preparing for the high-stakes tests. In other words, the perceptions of the test-takers can assist other test stakeholders, especially test developers, in comprehending various test-related factors. These factors include the test's usefulness, relevance to students' academic or professional goals, and alignment with the curriculum. Understanding test-takers' perceptions of tests is crucial, as it can provide insightful information to help test designers, instructors, and policymakers make improvements to assessments, testing practices, and practical policies.

Recent research in the testing field has focused on the perspectives of both test creators and test takers (Hu, 2022; Namfah, 2022), covering various dimensions, e.g., test takers' motivation, test performance, and language policy implementation. For instance, Hu (2022) surveyed 584 Chinese students regarding their perceptions of the high-stakes test, Zhongkao. The findings show that participants generally accepted computer-based testing and recognized its importance. Similarly, Wu and Lee (2017) found positive attitudes among Taiwanese university students toward an English proficiency test required for graduation, highlighting its value and motivational impact on language learning. Nevertheless, the relationship between the perceptions of university language policy and their test performance lacked significance. Moreover, Zhou and Yoshitomi (2019) investigated the correlation between perceptions of the language proficiency test, test-taking motivation, and test performance among 64 Japanese university students and found contradictory results. Their study shows a minimal correlation between test-takers' perceptions and motivation and language test performance.

Additionally, Imsa-ard (2020) explored the perceptions of 100 Thai upper secondary school teachers regarding the English test in the Ordinary National Educational Test (O-NET), and the study reveals that most teachers held negative attitudes towards the test, doubting its ability to accurately assess their students' real-life English language abilities. Furthermore, they believed that the test did not align with the national curriculum and should include other communicative skills beyond reading and grammar to enhance real-life communication abilities. Taweekae (2022) examined the perceptions and washback effects of the graduation requirement test among 256 Thai university students, and the study found that the test motivated students to focus on the goal and drive themselves to prepare for the graduation requirement test. However, both studies indicate that the high-stakes tests fail to reflect the actual outcomes of learning and teaching processes, leading to test-driven instruction rather than holistic language learning.

In the realm of high-stakes test design, development, and validation, several studies have investigated different aspects of the implementation and perception of CEFR-based proficiency tests. Waluyo et al. (2024), for instance, focused solely on interpreting quantitative results without including qualitative data from test takers. This approach has underscored the need for further qualitative investigation into this area. Meanwhile, Namfah (2022) and Phoolaikao

and Sukying (2021) explored test-takers' perceptions of the CEFR-based proficiency test. Both studies revealed that respondents exhibited a high level of awareness regarding the high-stakes test and CEFR implementation. However, they also highlighted a need for more information about the test and CEFR to adequately prepare test takers for their academic pursuits and future careers.

According to the review of literature on research regarding perceptions of high-stakes examinations, there is a lack of studies, particularly in the Thai context, concerning the test-takers' participation in designing the test. Consequently, this study primarily centers on the test-takers' perceptions of CMRU-TEP, which is the high-stakes language proficiency test for CMRU students, during the initial phases of test development and revision.

CMRU-TEP

Although the CEFR was initially adopted in Thai education, CMRU-TEP was specifically designed based on the Framework of Reference for English Education in Thailand (FRELE-TH), which was developed in 2017 based on CEFR (Nakanitanon, 2021). However, following the discontinuation of FRELE-TH, the development of new CMRU-TEP versions will now need to align with CEFR guidelines. Since its inception, CMRU-TEP has been administered to assess students' English language proficiency levels, ranging from A1 to C1. Specifically, it has been used to evaluate the English language proficiency of CMRU students in their final year prior to graduation. The primary objective of this assessment is to raise students' awareness of their English language proficiency levels, affording them ample time to prepare for the specific language requirements of their future workplaces.

The current version of CMRU-TEP has been administered for five years (2019-2023) and has been in the process of revision since the fourth quarter of 2023. The test consists of a total of 120 multiple-choice items, encompassing three main sections: listening, reading, and writing. Therefore, the assessment of students' scores remains somewhat approximate (Nakanitanon, 2021).

According to Nakanitanon (2021), the CMRU-TEP, developed by CMRU faculty, was validated by nine expert judges from the test construction team using the Yes/No Angoff method. The test demonstrated high content validity (0.95) and reliability (0.90). The time allocated to complete the test tasks is two hours and thirty minutes. The listening section comprises three parts with a total of 40 items. These items include picture-cued questions, short conversations, monologues with accompanying questions, and longer monologues featuring three or four questions. The reading section consists of 40 items divided into two parts. The first part includes various reading materials, such as signs, notices, announcements, maps, resumes, emails, brochures, job advertisements, and menus, totaling nine reading stimuli. The second part consists of six reading passages, each ranging from 100 to 300 words in length. The writing section includes 40 items designed as an indirect test with error detection, sentence completion, cloze test, and three to four passages (100 - 300 words long). The examination was administered by the CMRU Language Center staff.

The students' test scores were translated into CEFR levels as follows: Scores 0 - 21 correspond to No Level, scores 22 - 35 align with A1 Level, scores 36 - 56 equate to A2 Level, scores 57 - 79 indicate B1 Level, scores 80 - 104 signify B2 Level, and scores 105 - 120 represent C1 Level (Nakanitanon, 2021). Unfortunately, CMRU-TEP outcomes from 2019 to 2022 indicate that fewer than 16 percent of test-takers reached proficiency levels of B1 or B2, and fewer than two percent attained the C1 level (Chiang Mai Rajabhat University Language Center, 2023). These results do not align with the English proficiency standards outlined in university policies, reflecting impediments and challenges in the effective implementation of these policies, strategic plans, pedagogical practices.

Bachman and Palmer's model of test usefulness (1996)

According to Bachman and Palmer (1996), while designing and developing a test, usefulness is the principal quality, presenting the effectiveness and utilities of a language test. Bachman and Palmer's Model of Test Usefulness (1996) can be used as a framework to control the test's quality in educational and professional settings. The model consists of six test qualities:

- (1) Reliability is the consistency of language test scores, whether obtained on separate occasions, using varying measurement methods, or employing different assessments (Bachman & Palmer, 1996; Hughes & Hughes, 2020). When designing the test, the test designers attempt to minimize the inconsistencies of the measurement by taking the controllable factors affecting test performance under control. For instance, they ensure reliable scoring by providing precise guidelines to scorers through rubrics and by presenting clear test tasks to test-takers (Bachman & Palmer, 1996; Brown & Abeywickrama, 2018). Additionally, they actively encourage students to perform at their best during the test (Nitko & Brookhart, 2011).
- (2) Construct validity in language testing means the meaningfulness and appropriateness of justified interpretations drawn from the test scores, backed by substantial evidence to support the validity of the interpretations (Bachman & Palmer, 1996). For instance, if all items within a language test were intended to evaluate specific skills, yet there was a lack of empirical evidence to measure those skills of the test-takers, it would cast doubt on the construct validity of the test (Hughes & Hughes, 2020). Although the multiple-choice format often focuses on receptive skills (Jackson, 2022; Sato & Ikeda, 2015), incorporating CEFR descriptors into test specifications can enhance both the validity and reliability of the test (Al Lawati, 2023). Additionally, Brown and Abeywickrama (2018) suggest that assessments are more reliable and valid when they involve multiple measures rather than a single one. In line with this, construct validity can be enhanced by incorporating alternative assessments such as e-portfolios (Lam, 2023). Furthermore, Berry et al. (2019) recommend including a variety of assessments in teaching activities to improve overall evaluation effectiveness.
- (3) Authenticity is used to assess the degree to which the language tasks within a test align with real-world target language usage and relate to the test-takers' language requirements (Bachman & Palmer, 1996; Cheewasukthaworn, 2022). Tests lacking authenticity may not provide a precise interpretation and generalization of the test scores on an individual's

language abilities (Sangsuwan & Rukthong, 2023). To enhance the authenticity of a test, using authentic texts is crucial. Simonsen (2019) emphasizes that non-authentic texts are generally seen in a negative light; they are considered contrived, containing distorted and artificial language, and they lack sociocultural relevance.

- (4) Interactivity is used to evaluate the individual test-taker's language proficiency in real-world language usage, including language and topic knowledge, strategic competence, and affective responses to complete language test tasks (Bachman & Palmer, 1996; Cheewasukthaworn, 2022; Hughes & Hughes, 2020). According to Mattossi (2018) and Ying (2020), improving the interactivity of tests and test performance requires careful consideration of test-takers' topical knowledge, characteristics, and socio-cultural background during test content design. Additionally, Brown and Abeywickrama (2018) emphasize the importance of considering the length of passages when assessing language proficiency and performance, as it significantly influences readers' interaction with the text. Engaging with reading passages involves various cognitive processes, such as comprehension, interpretation, and integration of linguistic information, alongside psychological factors like memory, attention, and perception.
- (5) Impact of language testing encompasses both macro-level consequences within educational management systems and society, as well as micro-level effects on individual test-takers and users. This includes various aspects, such as test preparation, pedagogical methods, instructional materials, and classroom assessments (Bachman & Palmer, 1996; Brown & Abeywickrama, 2018). Moreover, Kuang (2020) examined empirical studies on the washback effect of English language tests on course content, materials, and activities in English language instruction. Kuang's study suggests that the washback of high-stakes tests can assist teachers in evaluating their impact and incorporating test content into the instructional process. The integration aims to align with course objectives and enhance both language development and test preparation, rather than solely focusing on exam-oriented language instruction (Rahman et al., 2021). Additionally, the studies conducted by Namfah (2022) and Wu and Lee (2017) highlight the positive impact of high-stakes test scores on students' perceptions toward the test and their continued English language studies. Furthermore, according to Mattossi (2018), raising awareness among language learners regarding their strengths and weaknesses throughout their learning journey is crucial for preparing them for high-stakes proficiency tests, as merely providing a numeric rating or score of the test is insufficient to motivate learners to address their learning difficulties and weaknesses.
- (6) Practicality relates to the balance between required resources and available resources and the effectiveness of available resource exploitation for test development (Bachman & Palmer, 1996; Sangsuwan & Rukthong, 2023). In the test development process, determining usefulness is a cyclical process where practicality considerations can influence decisions at any different stages, leading to a reevaluation and modification of some of the earlier specifications (Bachman & Palmer, 1996).

In conclusion, Bachman and Palmer's Model of Test Usefulness (1996) emphasizes the importance of language tests being valid, authentic, cognitively complex, and educationally impactful.

Language assessors can evaluate the overall usefulness of the language test in achieving its intended objectives by considering these six qualities. This model helps ensure that language tests provide test-takers, educators, and other stakeholders with relevant and useful insights. However, the CMRU-TEP was not designed according to this framework, which leads to concerns about the test's effectiveness. In this current study, these six characteristics were employed as the framework for gathering test-takers' perspectives on CMRU-TEP to assist in improving the test at a preliminary stage of test revision and development.

METHODOLOGY

Context of the study

This study is the initial stage of developing the CMRU-TEP according to the current Thailand's English language policy. However, the data used in this research were collected based on CMRU-TEP which was developed in 2017.

Research method

This study employs an explanatory sequential mixed-methods design, incorporating surveys and interviews, to explore the perceptions of CMRU-TEP among fourth-year students at Chiang Mai Rajabhat University (CMRU).

Population and sample

The population of this study comprises 2,696 fourth-year students (both male and female) at CMRU who took CMRU-TEP during the academic year 2023. To address the research questions, the respondents were divided into two groups: English majors and non-English majors. The major population included 278 students, while the non-major population comprised 2,418 students. The sample groups were composed of students who voluntarily responded to the questionnaire after completing CMRU-TEP. To enhance reliability, efforts were made to maximize sample sizes. Utilizing a quota sampling approach, a minimum of 30 percent of test-takers was included, resulting in 155 majors (representing 55.75% of the major population) and 882 non-major (representing 36.47% of the non-major population).

Data collection and data analysis

Data collection methods include the use of a questionnaire and interviews. The rating-scale questionnaire, ranging from 1 (strongly disagree) to 5 (strongly agree), consists of 25 questions. The first question requires respondents to identify whether they are an English major or a non-English major. The remaining 24 five-rating-scale questions assessing students' perceptions of CMRU-TEP based on the dimensions of reliability, construct validity, authenticity, interactivity, impact, and practicality (Bachman & Palmer, 1996). We developed the questionnaire and then validated it through three strategies. Firstly, we defined the construct required to be measured with the questionnaire to ensure alignment with the research

objectives. Secondly, two experts in English language teaching and evaluation reviewed the survey items to ensure they comprehensively covered the construct and that the language used was accurate and appropriate. Expert 1 recommended changing the questions from interrogative to declarative form to better facilitate self-evaluation by the students. Consequently, all questions were revised to "I" statements, such as "I believe that the CMRU-TEP's scores reliably measure English language abilities". Expert 2 suggested making the language more concise to enhance student understanding. Therefore, longer questions were shortened.

The questionnaire was translated into Thai to accommodate students and obtain more accurate answers about their perceptions of the test. An English-to-Thai translation expert then verified the equivalence of the translated version with the original text. It was determined that the translation achieved: 1) linguistic equivalence, reflecting the same meaning as the original English text; 2) semantic equivalence, conveying the same meanings; and 3) conceptual equivalence, ensuring that the concepts tested by the questionnaire were understood similarly in both cultures.

Then, the questionnaire underwent pilot testing with a sample of 12 respondents, fourth-year students who were not included in the study. They completed the questionnaire in about seven minutes. The students noted that they could have finished it more quickly if it had been available in only one language. Because it was provided in both English and Thai, they ended up reading both versions, which slowed them down. To address this, the questionnaire was later divided into two separate versions, allowing students to choose the one they prefer. The language used in the instructions was clear; however, the last question, which inquired whether the CMRU-TEP's cost and registration process aligned with students' financial resources and convenience, confused two students. Therefore, the question was revised to focus on a specific aspect, now asking, "I think the CMRU-TEP's cost aligns with my financial resources".

The development of the semi-structured interview form involved a meticulous process to ensure its effectiveness. Initially, the form comprised six main questions in Thai, all derived from the theoretical framework. These questions were first presented to three experts to gather their feedback and suggestions for improvement. They agreed with the six questions as they were, suggesting no changes. After that a pilot test was conducted with five fourth-year students who were not part of the primary data collection group. The feedback from this trial indicated the need for additional sub-questions to enhance clarity, leading to further refinements of each main question. These adjustments were made before the actual data collection to guarantee the accuracy, suitability, and comprehensibility of the questions. After these refinements, the updated data collection tools were submitted to the Institutional Review Boards (IRB) for ethical approval.

After completing the 120-item CMRU test, each student was asked by the test administrator to fill out the questionnaire provided via a QR code. To identify which quantitative results required additional clarification regarding students' perceptions of CMRU-TEP and to delve deeper into their experiences and opinions, semi-structured interviews were conducted. Interviews were held with five English majors and ten non-English majors, including five from the Science fields and five from the Arts fields. Interviewees were selected based on the

criterion that they were among the top five respondents who scored the lowest on their questionnaire responses.

For the qualitative data obtained from interviews, we conducted a concept-driven analysis. The data from the questionnaire were analyzed quantitatively to find the percentage, mean, standard deviation, and independent samples t-test. The scores were interpreted regarding the following criteria:

- 4.51 - 5.00 = The students perceived that CMRU-TEP was highly useful.
- 3.51 - 4.50 = The students perceived that CMRU-TEP was useful.
- 2.51 - 3.50 = The students perceived that CMRU-TEP was moderately useful.
- 1.51 - 2.50 = The students perceived that CMRU-TEP was slightly useful.
- 1.00 - 1.50 = The students perceived that CMRU-TEP was not useful.

RESULTS AND DISCUSSION

To address the first research objective regarding the perceptions of CMRU-TEP among fourth-year students, both English majors and non-English majors, the survey and interview results yield the following insights.

Table 1
Overall differences of English majors and non-English majors' perceptions

| Topic | Majors | | Non-Majors | | t | sig (2-tailed) |
|-------------------------|-----------|------|------------|------|------|-------------------|
| | \bar{X} | S.D. | \bar{X} | S.D. | | |
| Perceptions of CMRU-TEP | 3.73 | 0.95 | 3.68 | 0.90 | 1.73 | .09* |

* $p < .05$

Table 1 shows that English majors and non-English majors perceived CMRU-TEP with average scores of 3.73 and 3.68, respectively. Additionally, their perceptions do not differ at the .05 level, suggesting that both groups of students perceived CMRU-TEP as useful, aligning with Bachman and Palmer's (1996) six qualities of test usefulness.

Table 2
Mean scores, standard deviations (S.D.), and difference in mean scores for six qualities of test usefulness

| Qualities of Test Usefulness | Majors | | | Non-Majors | | | Difference in Mean Scores |
|------------------------------|-------------|-------------|---------------|-------------|-------------|---------------|---------------------------|
| | \bar{X} | S.D. | Meaning | \bar{X} | S.D. | Meaning | |
| 1. Reliability | 3.58 | 1.00 | Useful | 3.62 | 0.92 | Useful | 0.04 |
| 2. Construct validity | 3.70 | 0.98 | Useful | 3.70 | 0.89 | Useful | 0.00 |
| 3. Authenticity | 3.66 | 0.92 | Useful | 3.63 | 0.90 | Useful | 0.03 |
| 4. Interactiveness | 3.87 | 0.90 | Useful | 3.71 | 8.88 | Useful | 0.16 |
| 5. Impact | 3.81 | 0.92 | Useful | 3.71 | 0.88 | Useful | 0.10 |
| 6. Practicality | 3.74 | 0.93 | Useful | 3.68 | 0.92 | Useful | 0.06 |
| Average | 3.73 | 0.95 | Useful | 3.68 | 0.90 | Useful | 0.05 |

Table 2 indicates that among the six qualities, “Interactiveness” exhibited the highest difference in mean scores (0.16), followed by “Impact” (0.10) and “Practicality” (0.06).

To identify areas for improvement in CMRU-TEP, highlighted by students who expressed disagreement with the six qualities of test usefulness in survey results, interviews were conducted with 15 selected participants who met the recruitment criterion. The results for each quality are presented below, starting with the students’ perspectives, followed by strategies to enhance CMRU-TEP, and concluding with support mechanisms to improve test performance.

Quality 1: Reliability

The findings from both groups of interviewees indicate that the highest disagreement regarding reliability stems from the test administration process (46.67%), concerns about potential guessing (40.00%), and a disparity between reported scores and students’ perceived proficiency (13.33%).

Among non-majors, the majority (50.00%) expressed doubts about the reliability of the reported scores due to the test administrators reporting scores twice, possibly due to calculation errors. Furthermore, the lengthy reading passages prompted them to guess answers without thoroughly reading through the passage.

“I think the reading was too long, so students were likely to guess rather than finish reading. Additionally, scores were announced twice even though there was only one exam. The scores from the first and second rounds were not the same, and some who did not take the exam still received scores.” (Interviewee 1, non-major)

A considerable number of students (46.67%, comprising 13.33% of non-majors and 33.33% of majors) received CMRU-TEP scores lower than their expected results.

“I once took the TOEIC exam during high school and scored 789, which corresponds to the C2 level of the CEFR. However, I am currently a fourth-year university student. Why did I only score at the B1 level in the CMRU-TEP exam?” (Interviewee 10, major)

“I scored lower than I expected. Some of my friends who were weaker than me scored higher, while others who were better scored lower. I wasn’t very confident in the exam scoring method.” (Interviewee 12, non-major)

The research findings suggest several potential concerns regarding the inconsistencies of CMRU-TEP attributed to the test administration process. Additionally, there was a lack of transparency or communication regarding how scores were determined, leading students to speculate about the scoring process. As suggested by Hughes and Hughes (2020, p. 71), a handbook that includes details of the test, related materials, and the test administration process is needed for the test administration. Therefore, the test center should maintain detailed records of the test administration process and then develop a comprehensive administration manual that provides clear instructions for test administrators. They should

ensure that test administration follows standardized procedures. Training sessions should also be administered to familiarize test administrators with the test content, format, and administration protocols and ensure consistent implementation.

Test administrators also need to implement strict test security measures to prevent cheating. If applicable, they should randomize the order of test items to minimize the impact of order effects on test scores. As suggested by the interviewees, changing the test to a computer-based format could further enhance randomization, helping to control any systematic biases introduced by the order of presentation. Additionally, conducting regular reviews of CMRU-TEP to identify and address any potential issues with the test's reliability can lead to continuous improvements in the test design and administration.

Providing ample opportunities for practice and exposure to test-like situations can enhance the consistency of CMRU-TEP scores by reducing student anxiety and improving their test-taking skills. Integrating test training into English language courses ensures a comprehensive understanding of specific aspects (Phoolaikao & Sukying, 2021). Both English instructors and subject matter instructors should actively motivate students to achieve their best performance during the test (Nitko & Brookhart, 2011). Crucially, English instructors should utilize the CEFR as a framework for shaping classroom practices and assessments. This comprehensive approach ensures a cohesive and supportive environment for students undertaking CMRU-TEP, fostering a more effective and integrated learning experience.

Quality 2: Construct validity

Both groups of interviewees (100%) reported that CMRU-TEP covered three areas of language knowledge: listening, reading, and grammatical knowledge. Notably, it omitted the evaluation of speaking and writing abilities. Therefore, the attributes of the test tasks in CMRU-TEP should not be construed as reliable indicators of students' proficiency in English.

“I believe that a cloze test cannot serve as a replacement for a writing proficiency test. Instead, I view it as a measure of reading skills and the ability to infer the meaning of vocabulary from contextual clues.” (Interviewee 6, major)

“The exam lacked speaking skills and had few everyday life skills.” (Interviewee 1, non-major)

“The writing section should have allowed students to write, such as short answers and essays.” (Interviewee 13, non-major)

CMRU-TEP was found to have a predominant focus on evaluating receptive skills, specifically listening and reading, as well as grammatical knowledge. This observation aligns with the findings of Jackson (2022) and Sato and Ikeda (2015). Converting the CEFR descriptors into specifications can ease the process of crafting items effectively and enhance the tests' validity and reliability (Al Lawati, 2023). However, limitations in university resources may have resulted in a narrower inclusion of speaking and writing assessments. Designing, administering, and

evaluating speaking and writing assessments can be resource-intensive, potentially imposing constraints on the university in terms of time, personnel, or budget. To address both the university's limitations and the need for robust testing, improving the quality of CMRU-TEP could include more measures, as suggested by Brown and Abeywickrama (2018). They can maintain e-portfolios (Lam, 2023) with records of speaking and writing assessments, contributing to the overall English proficiency test scores.

To improve test performance, English instructors need the ability to actively engage in diverse teaching, learning, and assessment methods simultaneously (Berry et al., 2019). Language planners should communicate their goals to various stakeholders, including English programs, Language Centers, non-major programs, and students. Instead of relying solely on a proficiency test at the end of a four-year program, classroom instructors should incorporate ongoing assessments before reaching conclusions about students' English proficiency.

Quality 3: Authenticity

Most of the students (46.67%, comprising 26.67% of non-majors and 20.00% of majors) agreed that CMRU-TEP did not align with the content covered in the English courses offered by their programs. Despite non-majors completing two compulsory fundamental English courses in their first or second year of study, majors engaged in at least 25 English courses. Majors stated that the test emphasized everyday tasks, which they found less challenging compared to the lessons and materials covered in their English courses. They expressed a need for the test to be more academically oriented and connected to professional contexts than it was.

“I studied more than what was covered in the exam. I think the exam focused on everyday life. I would have preferred it to emphasize academic skills necessary for future careers.” (Interviewee 10, major)

Conversely, non-majors expressed that the test emphasized academic content, posing difficulty for them. They believed that the test required more skills than what was taught in their English courses.

“I took only two English courses in my first year. During the exam, I used limited skills from what I had learned. I relied more on guessing. The reading passages in the exam were much longer and more difficult than the ones I had studied, which were easy and short.” (Interviewee 1, non-major)

The substantial difference in English exposure—over 1,000 hours for majors versus fewer than 100 for non-majors—likely shapes their contrasting perceptions on the CMRU-TEP's difficulty and authenticity. English majors, with greater exposure and higher expectations, often perceive the test as lacking academic rigor due to its focus on communicative rather than academic language. However, as Bachman and Palmer (1996) and Cheewasukthaworn (2022) emphasize, authenticity in language testing lies in reflecting real-life language use. In this light, CMRU-TEP's communicative focus aligns well with authentic assessment principles. To address non-majors' concerns about test difficulty, CMRU test developers should first identify key features of

students' actual English use and the university's language requirements, then design or select tasks that align with these needs.

However, specifying the domain for a general proficiency test is challenging due to variations in the target language based on the field of study (Zheng & Iseni, 2017). It would be beneficial for CMRU-TEP to be designed to reflect the focus of communicative skills and content emphasized in the university's English courses. The test items, texts, and tasks should align with real-life communicative situations (Giraldo, 2019) and mirror the cognitive processes of test-takers in both testing and non-testing scenarios (Hasrol et al., 2022). Furthermore, it is recommended that CMRU-TEP incorporates authentic texts that are both time-dependent and context-based (Simonsen, 2019).

Potential misalignment between the content and the university's English course curriculum may cause students to perceive a disconnect between classroom instruction and the assessment in CMRU-TEP. To address this, the test should align its content with language usage in the target language use domain, curricula, and textbooks (Bachman, 1991; Hasrol et al., 2022). Incorporating authentic academic texts and situations in classroom teaching (Glisan & Donato, 2017) can further bridge the gap between theoretical learning and practical application, fostering critical thinking.

University policymakers need to clearly communicate CMRU-TEP expectations to all stakeholders and offer sufficient support. Additionally, collaboration between English programs and subject matter programs is crucial for designing contextualized English courses tailored to non-majors.

Quality 4: Interactiveness

Seventy percent of non-majors found the reading passages excessively long, complex, and difficult, leading to feelings of boredom and minimal engagement. Consequently, they employed a limited range of language skills and experienced lower levels of cognitive and emotional involvement.

"I think the reading test had too much content, and it was difficult. When working on it, I didn't know when it would be finished, making my brain tired. As a result, I had to guess a lot." (Interviewee 13, non-major)

Majors reported that not only were the passages too long and boring, but their content was also not challenging. Moreover, the vocabulary was simpler than what they learned in the classroom, resulting in a lack of the use of metacognitive strategies, topical knowledge, and affective responses.

"I didn't feel engaged with the reading test at all because the vocabulary was too simple, making it boring. My mind didn't stay focused, and I had to read multiple times. If more interesting vocabulary were used, I would have found it more enjoyable to read." (Interviewee 7, major)

To address these issues, test developers should consider the variability in test tasks across CEFR levels, ensuring inclusivity for diverse proficiency levels. CMRU, as a tertiary institution dedicated to fostering inclusivity and expanding opportunities for the local community, admits high school students, many of whom possess limited proficiency in English. Therefore, it is essential for test designers to prioritize the allocation of test tasks, ensuring that the test includes familiar and localized topics suitable for diverse English language proficiency levels and sociocultural backgrounds (Matoussi, 2018; Ying, 2020). Moreover, given that CMRU-TEP primarily employs multiple-choice formats, introducing authentic speaking and writing tasks could enhance test-taker engagement.

Moreover, the test tasks should accommodate a range of language proficiencies without being excessively simple or difficult compared to the expected CEFR levels. This balance is crucial for accurate and effective test performance, which requires both cognitive skill development and affective enhancement (Bachman & Palmer, 1996; Hughes & Hughes, 2020).

Additionally, considering the length and cognitive complexity of passages should be taken into account. Brown and Abeywickrama (2018) suggest that the quantity of paragraphs per page influences how readers engage with text psychologically. Genres such as anecdotes, short narratives, descriptions, excerpts from longer texts, announcements, directions, and similar formats are particularly conducive to interactive reading experiences. Therefore, to enhance the test's interactivity, test developers might consider shortening lengthy tasks or divide them into smaller, more manageable sections based on relevant topics and contexts mentioned in the CEFR. Shortening texts could serve the dual purpose of meeting students' needs and aligning with the established language proficiency standards outlined in the CEFR, which provides a framework for assessing language proficiency across various levels, from A1 (beginner) to C2 (proficient).

In addition to improving the test performance, recognizing individual differences in language strengths and weaknesses is important to support CMRU students in improving their test performance. English language instructors should identify students' language weaknesses during the courses to motivate them to address their language skill gaps before undertaking the high-stakes test (Matoussi, 2018). Moreover, familiarity with the format, structure, and expectations of language tests can benefit test-takers, enabling them to engage more effectively with the tasks presented to them (Moulden, 2018).

By pinpointing students' weaknesses and providing targeted support, guiding students through practice and test preparation may possibly alleviate their feelings of boredom or anxiety during the exam and help them develop the language skills needed for success in both high-stakes tests and real-life communication scenarios. When learners become aware of their language learning challenges, they will try to overcome the challenges by employing cognitive and metacognitive strategies, resulting in a positive impact known as the washback effect on language learning.

However, concerns raised by Ying (2020) highlight a discrepancy between the effectiveness of communication teaching methods and the test-oriented nature of localized language policies.

To address this potential gap in the context of the study, CMRU English language instructors, both for non-English majors and English majors, along with test developers, should engage in discussions about course learning objectives, CMRU-TEP goals, and university language policy expectations. This collaborative effort fosters mutual understanding and enables instructors to design language courses that effectively serve course objectives, align with CMRU-TEP's goals, and meet university language policy expectations. Additionally, as Kuang (2020) suggests, English language teachers can assess if a test will benefit their teaching and integrate its content into the instructional process, aligning with course goals and aiding student language development and test readiness. Furthermore, guiding students through practice and test preparation is about empowering them with skills and strategies to succeed in testing contexts, fostering confidence, reducing anxiety, and improving overall test-taking abilities.

For CMRU-TEP test developers, it is essential to consider the topics incorporating the content covered in English language courses and aligning with the CEFR. By doing so, test-takers may experience reduced anxiety and improved interaction with the test, as they will have acquired relevant topic knowledge from their language lessons that align with the test content.

Quality 5: Impact

All interviewees (100%) revealed that CMRU-TEP did not influence their motivation to further develop their English language skills. Majors (26.67%) reported that the test was too easy to encourage further development, while non-majors stated that the test scores had no impact on their future careers (26.67%), and a significant portion did not prepare for the test (40.00%).

“There is no motivation to improve oneself because the test is too easy and lacks challenge. It has no impact on language use and my further self-development in language learning.” (Interviewee 6, major)

“I am not an English major, so the test scores do not impact my future work, and they do not contribute to my motivation to develop English language skills since I do not need to use them.” (Interviewee 12, non-major)

“I don’t know why we’re taking the test and what the test scores will affect. There’s no preparation beforehand.” (Interviewee 14, non-major)

All participants in this study reported that the test failed to motivate or prompt them adequately to prepare or advance their English skills. This aligns with the findings of Taweekae (2022). The deficiency in the test’s “Impact” quality attributed to its scores lacking influence on their future careers and academic pursuits. Thus, recognizing the test’s significance for their career prospects could enhance students’ motivation and involvement (Namfah, 2022).

To address this issue, it is important to consider the alignment of language learning objectives, English course content and CMRU-TEP with the CEFR standards. This ensures that the test scores reflect the expected language skills of the students across different proficiency levels, while also raising awareness of the test’s impact on their language learning progress.

As suggested by Kuang (2020), integrating test content and topics into teaching practices can significantly enhance student language development and prepare them for the test.

Moreover, for non-English majors who may lose interest in using English language after completing their courses and the proficiency test, it is essential to implement strategies that go beyond the test itself and encompass broader aspects of language learning and support within the university environment. The university can foster a language-rich environment by providing opportunities for English language use in various contexts beyond the test, such as extracurricular activities, language exchange programs, or immersion experiences that encourage students to continue using English in meaningful ways outside of the classroom. Then, the students can perceive the relevance of their English language studies to their overall academic and personal development, thereby increasing their motivation to continue using English beyond the test.

To ensure effective policy implementation and support, collaboration among stakeholders is vital. For instance, the Language Center should coordinate with the registrar to communicate the test schedule, and subsequently, the registrar should notify individual fourth-year students. The library should provide ample resources for English learning, while the language center should offer test preparation resources, workshops, or courses specifically addressing the skills and content covered in CMRU-TEP.

Quality 6: Practicality

The interview findings exhibited that most students (80.00%, comprising 53.33% of non-majors and 26.67% of majors) highlighted the ill-preparedness of test administrators. Non-majors specifically noted an insufficient number of proctors (33.33%), while majors emphasized the tardiness of proctors (20.00%), both in terms of being late for duty and exhibiting slow responses to students' inquiries.

“There were only a few proctors, around 2 per group of approximately 100-150 students. Ideally, the number of examinees should be divided into groups of 30-40 people per room, with at least 2 proctors per room” (Interviewee 15, non-major)

“The proctors were not punctual. The exam was scheduled to start at 1:00 p.m., but the proctors arrived at 1:40 p.m. They even rushed back to get some test equipment. However, when the exam time ended, there was no extension to compensate for the lost time. The seating arrangements for the exam were also inappropriate, and there was loud noise every time someone adjusted their position.” (Interviewee 6, major)

Many students (60.00%) also mentioned problems related to material resources. Non-majors (26.67%) expressed concerns about factors facilitating cheating opportunities, such as the reuse of test papers and seating arrangements, while majors (13.33%) highlighted equipment malfunctions, particularly issues with the sound system. Additionally, some non-majors (20.00%) expressed concerns about insufficient time for completing the test.

"There was the use of old test papers where other students had already selected answers, making me hesitate in my own choices. I would prefer a switch to online exams and an extension of the exam duration to 3 hours." (Interviewee 12, non-major)

"The exam equipment, particularly the chairs, was not properly prepared. They produced noise when moved, leading to distractions and disrupting concentration during the exam." (Interviewee 6, major)

"The audio system was malfunctioning. It was a hindrance during the listening test." (Interviewee 7, major)

"The number of exam questions was 120, but the allotted time for the exam was short. The exam duration should be extended." (Interviewee 13, non-major)

According to the interview, a significant concern raised by students is the reuse of paper-and-pencil tests, prompting the consideration of computer-based testing as a potential alternative. This shift could mitigate cheating possibilities, cut printing costs, and enable better allocation of testing resources. Regarding test time allocation, test developers should account for the relationship between the number of test items, task complexity, and time allocation. Thus, conducting pilot studies across different sample groups can ensure an appropriate test completion time.

Moreover, an impractical test is characterized by high costs, time-consuming processes, and unclear procedures and administration (Brown & Abeywickrama, 2018). Therefore, enhancing students' test performance and the test's practicality requires careful consideration of CMRU-TEP arrangements, pertaining to human resources, material resources, and time allocation, from the preparation stage to the score announcement stage.

The number of test items can be reduced from 120 to 100 while still encompassing the measurable skills necessary for the university's graduation requirements. Functional testing and grading equipment, as well as a suitable testing environment, are necessary. This ensures that students can effectively exhibit their knowledge and language proficiency without being disturbed by external factors. The availability and accessibility of resources for test administration and implementation must be considered. Therefore, the university should prioritize practical testing equipment and conducive environments to bolster students' test performance, thereby minimizing factors that might impact their test results. Furthermore, those who administer the test should be adequately trained in how to maintain the test's practicality and fairness, preventing resource wastage and ensuring effective, efficient test administration for the ultimate benefit of test takers.

CONCLUSION AND RECOMMENDATIONS

Given the significance of test takers' voices in test design and the challenge of developing a general English proficiency test for diverse fields of study, this study primarily focuses on

gathering the perceptions of majors and non-majors regarding a high-stakes language proficiency test, namely CMRU-TEP, in the initial phases of test development and revision.

While emphasizing reliability and validity is crucial for a high-stakes test (Bachman & Palmer, 1996), a primary concern of CMRU-TEP's reliability was caused by the test administration process. The resolution of this concern requires a detailed handbook for the test administration and well-trained administrators. Limitations in resources inevitably led CMRU-TEP to primarily focus on receptive skills and grammatical knowledge. We propose that a portfolio assessment should be employed to address the limitations imposed by the university's resource constraints. In addition, for a university's in-house proficiency test for students with low English proficiency, the primary goal should be to measure communicative competence, ensuring usefulness for both majors and non-English majors.

Improving test performance for both majors and non-majors involves identifying individual language weaknesses and focusing on those areas before the test. Offering guidance in test preparation, familiarizing students with formats, and promoting awareness of the test's importance are crucial. Besides, prioritizing practical language policies, aligning test objectives with instruction, and collaborating among developers, instructors, and curriculum planners are vital. Clear communication of test expectations, practice opportunities, and efforts to reduce test-related anxiety on non-majors are crucial. Additionally, creating understanding among policymakers, implementers, and stakeholders during policy launches is also key. Adding prestige to test scores and the test itself could boost stakeholder awareness. Moreover, ensuring trained proctors, responsive staff, and prepared administrators enhances test practicality. Finally, providing functional equipment and conducive testing environments could minimize distractions for better test performance.

Although this study offers insightful perspectives on the problems of an in-house high-stakes test from the viewpoints of the test-takers, it was limited to CMRU students. Therefore, the findings cannot be generalized to all Thai students or to other EFL contexts. Moreover, the research focused solely on the perspectives of test takers regarding test usefulness. Future studies could broaden this understanding by exploring the views of test administrators, proctors, and other stakeholders. Additionally, further research might examine the test's effects at both micro and macro levels. For instance, studies could investigate the test's washback on English language teaching and learning. Researchers could also explore how the test influences policy and planning in English language education.

ACKNOWLEDGMENTS

The authors would like to acknowledge the use of AI-assisted language editing, in accordance with the journal's ethical guidelines, for refining the language to enhance the accuracy, coherence and readability of this manuscript.

THE AUTHORS

Panachanok Chanwaiwit, PhD, is an assistant professor at Chiang Mai Rajabhat University, Thailand, with 20 years of teaching experience across primary, secondary, and higher education. Her research is centered on English language teaching, teacher education, and technology-enhanced learning.

panachanok_cha@cmru.ac.th

Lalida Wiboonwachara, PhD, is a lecturer at Chiang Mai Rajabhat University, Thailand. She attained her PhD in English for Professional Development at the School of Liberal Arts, Mae Fah Luang University, Thailand. Her research interests mainly lie in writing instruction, self-regulation, and critical thinking.

lalida_wib@cmru.ac.th

REFERENCES

Al Lawati, Z. A. (2023). Investigating the characteristics of language test specifications and item writer guidelines, and their effect on item development: A mixed-method case study. *Language Testing in Asia*, 13, Article 21. <https://doi.org/10.1186/s40468-023-00233-5>

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671–704. <http://doi.org/10.2307/3587082>

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113–123. <https://doi.org/10.1093/elt/ccy055>

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.

Chartrakul, K., & Damnet, A. (2021). Role of the CEFR and English teaching in Thailand: A case study of Rajabhat universities. *Advances in Language and Literary Studies*, 12(2), 82–89. <http://dx.doi.org/10.7575/aiac.all.v.12n.2.p.82>

Cheewasukthaworn, K. (2022). Developing a standardized English proficiency test in alignment with the CEFR. *PASAA Journal*, 63, 66–92. <https://doi.org/10.58837/CHULA.PASAA.63.1.3>

Chiang Mai Rajabhat University. (2023, June 1). *Announcement of Chiang Mai Rajabhat University on English language proficiency enhancement for the students*. <http://www.general.cmru.ac.th/infocenter/condec/ConCmdFiles/2566/582566.pdf>

Chiang Mai Rajabhat University Language Center. (2023). *CMRU-TEP score report for final-year undergraduate students (2019-2022)* [Unpublished manuscript]. Chiang Mai Rajabhat University Language Center, Chaing Mai Rajabhat University.

Commission on Higher Education Standards. (2024, January 17). *Announcement of the Office of Higher Education Commission on the English language proficiency enhancement national policy A.D. 2024*. <https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625-2567>

Glisan, E. W., & Donato, R. (2017). *Enacting the work of language instruction: High-level teaching practices*. American Council on the Teaching of Foreign Languages.

Giraldo, F. (2019). Designing language assessments in context: Theoretical, technical, and institutional considerations. *HOW: A Colombian Journal for Teachers of English*, 26(2), 123–143. <https://doi.org/10.19183/how.26.2.512>

Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(2022), Article 100023. <https://doi.org/10.1016/j.rmal.2022.100023>

Hu, H. (2022). Computer-delivered English listening and speaking test in Zhongkao: Test-taker perception, motivation and performance. In F. Uslu (Ed.), *Proceedings of SOCIOINT 2022 - 9th International Conference on Education & Education of Social Sciences* (pp. 59–75). International Organization Center of Academic Research. <https://doi.org/10.46529/socioint.202209>

Hughes, A., & Hughes, J. (2020). *Testing for language teachers* (3rd ed.). Cambridge University Press.

Imsa-ard, P. (2020). Voices from Thai EFL teachers: Perceptions and beliefs towards the English test in the national examination in Thailand. *LEARN Journal: Language Education and Acquisition Research Network*, 13(2), 269–287. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/243713>

Jackson, S. (2022). Student and teacher perceptions of an English language test for university admission: Understanding its face validity. *Journal of Language and Culture*, 41(2), 171–220. <https://so03.tci-thaijo.org/index.php/JLC/article/view/265593>

Jantadej, K. (2021). English as an international language teaching and perceptions: A case study of Thai tertiary English language teachers. *International Journal of Progressive Education*, 17(6), 93–114. <https://doi.org/10.29329/ijpe.2021.382.7>

Kasschau, R. A. (2003). *Understanding psychology*. McGraw-Hill.

Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching*, 13(9), 10–17. <https://doi.org/10.5539/elt.v13n9p10>

Lam, R. (2023). E-portfolios: What we know, what we don't, and what we need to know. *RELC Journal: A Journal of Language Teaching and Research*, 54(1), 208–215. <http://doi.org/10.1177/0033688220974102>

Mattoussi, Y. (2018). Testing usefulness of reading comprehension exams among first year students of English at the tertiary level in Tunisia. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 265–288). Springer. https://doi.org/10.1007/978-3-319-62884-4_13

Moulden, M. (2018, July 23). *Test interactiveness: A vital quality*. <https://tesolsolutions.wordpress.com/2018/07/23/test-interactiveness-a-vital-quality/>

Nakanitanon, P. (2021). Linking an English proficiency test to the CEFR: Setting valid cut scores. *New English Teacher*, 15(1), 53–64. <http://www.assumptionjournal.au.edu/index.php/newEnglishTeacher/article/view/4878>

Namfah, N. (2022). Perception towards the CEFR-based English proficiency test: A case of the students-teacher in the teacher for local development project at one university in Thailand. *Journal of Education Burapha University*, 33(3), 16–38. <https://so02.tci-thaijo.org/index.php/edubuu/article/view/256515>

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Pearson.

Office of the National Economic and Social Development Council. (2022, October). *The thirteenth national economic and social development (2023-2027)*. https://www.nesdc.go.th/ewt_news.php?nid=13651&filename=develop_issue

Phoolaikao, W., & Sukying, A. (2021). Insights into CEFR and its implementation through the lens of preservice English teachers in Thailand. *English Language Teaching*, 14(6), 25–35. <https://doi.org/10.5539/elt.v14n6p25>

Rahman, K. A., Seraj, P., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11(12), 1–23. <https://doi.org/10.1186/s40468-021-00129-2>

Rohman, K. A., Budiana, H., & Hartini, N. (2019). The students' perceptions on a test of English proficiency as graduation requirement. *LLT Journal: A Journal on Language and Language Teaching*, 22(2), 171–181. <https://doi.org/10.24071/llt.2019.220204>

Sangsuwan, W., & Rukthong, A. (2023). Test-takers' performances on and perceptions of two different modes of online speaking tests. *LEARN Journal: Language Education and Acquisition Research Network*, 16(2), 168–183. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/266940>

Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on students learning. *Language Testing in Asia*, 5(1), Article 10. <https://doi.org/10.1186/s40468-015-0019-z>

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2009). *Psychology*. Worth Publishers.

Simonsen, R. (2019). An analysis of the problematic discourse surrounding authentic texts. *Hispania*, 102(2), 245–258. <http://doi.org/10.1353/hpn.2019.0045>

Taweekaew, K. (2022, May 8). *Washback effects on implementing English proficiency tests as a graduation requirement at Thailand National Sports University (TNSU)* [Paper presentation]. The 9th NEU National and 7th International Conference, Northeastern University, Khon Kaen, Thailand. https://idcneu.com/neunic22/file_article/ic/neunic2022-ic-07.pdf?utm_source=chatgpt.com

van de Watering, G., Gijbels, D., Dochy, F., & van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education*, 56(6), 645–658. <http://www.jstor.org/stable/40269095>

Waluyo, B. (2019). Thai first-year university students' English proficiency on CEFR levels: A case study of Walailak university, Thailand. *New English Teacher*, 13(2), 51–71. <http://www.assumptionjournal.au.edu/index.php/newEnglishTeacher/article/view/3651>

Waluyo, B., Zahab, B., & Ruangsung, L. (2024). Language assessment at a Thai university: A CEFR-based test of English proficiency development. *rELFlections*, 31(1), 25–47. <https://doi.org/10.61508/refl.v31i1.270418>

Wittayasin, S. (2017). Education challenges to Thailand 4.0. *International Journal of Integrated Education and Development*, 2(2), 29–35. <https://so02.tci-thaijo.org/index.php/ijied/article/view/131930>

Wu, J., & Lee, M. (2017). The relationship between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities. *Language Testing in Asia*, 7(1), 1–21. <https://doi.org/10.1186/s40468-017-0041-4>

Ying, B. (2020). The relationship of take takers' learning motivation, attitudes towards the actual test use and test performance of the College English Test in China. *Language Testing in Asia*, 10(10), 1–18. <https://doi.org/10.1186/s40468-020-00108-z>

Zheng, Y., & Iseni, A. (2017). Authenticity in language testing. *Anglisticum Journal*, 6(8), 9–14. <https://media.neliti.com/media/publications/570733-authenticity-in-language-testing-61a6682a.pdf>

Zhou, Y., & Yoshitomi, A. (2019). Test-taker perception of and test performance on computer-delivered speaking tests: The mediational role of test-taking motivation. *Language Testing in Asia*, 9, 1–19. <https://doi.org/10.1186/s40468-019-0086-7>