Developing instrument to measure critical thinking abilities of Thai undergraduate students

Tanyapon Phongphio[a], Beatriz Gallo Cordoba[b]

[a]Chiang Mai University, Thailand, [b]Victoria University, Australia

ABSTRACT

Thailand's National Qualifications Framework emphasizes critical thinking in higher education. However, assessing Thai students' critical thinking skills is challenging without a Thai-language version of the widely recognized Watson-Glaser Critical Thinking Appraisal (WGCTA). A culturally adapted assessment is needed to eliminate language barriers. This study aimed to develop a Thai critical thinking test based on the WGCTA format to assess undergraduate students' skills. To address this issue, the study adopted the design-based research (DBR) framework. In stage 1 of the DBR process, analysis revealed the need for a critical thinking test for Thai undergraduates. Stage 2 involved developing a Thai-translated version, and stage 3 piloted the prototype with 52 participants. This article presents the pilot study results. The findings showed low reliability of the instrument, limiting its effectiveness as a measurement tool. The factors included the high degree of homogeneity among the test group, the inappropriate difficulty level of the test items, and hence inefficiencies in the test's ability to distinguish between high-performing and low-performing test takers. These insights are critical for informing subsequent revisions to ensure that the instrument becomes a reliable and valid tool for assessing Thai students' critical thinking.

KEYWORDS

assessment of critical thinking, critical thinking tests, assessment tool development, critical thinking development

CORRESPONDENCE

Tanyapon Phongphio (tanyapon.p@cmu.ac.th)

INTRODUCTION

Critical thinking is recognized as a crucial learning outcome for students in higher education, as outlined in the National Qualifications Framework (Thailand NQF) (Office of the Education Council, 2017). The emphasis on fostering critical thinking skills is not only an educational priority but also a national developmental goal. It aligns closely with the objectives of the Twelfth National Economic and Social Development Plan (2017–2021), which aims to prepare a skilled workforce capable of meeting the demands of the 21st century. The integration of critical thinking into educational objectives reflects a belief that equipping students with the ability to think critically and make reasoned judgments will contribute to shaping a society of informed, rational, and responsible citizens in the future. The National Qualifications Framework for Higher Education in Thailand (Office of the Higher Education Commission, 2006) explicitly outlines the expectations for students' critical thinking abilities. It was already articulated that a well-educated graduate from a bachelor degree programme should possess cognitive skills and be able to:

Carry out investigations, comprehend and evaluate new information, concepts and evidence from a range of sources, and apply conclusions to a wide range of issues and problems without external guidance. Is able to investigate complex problems and recommend creative and innovative solutions taking account of relevant theoretical knowledge and practical experience and the consequences of decisions made. Can apply these skills and insights in professional and academic contexts relevant to the field of study undertaken. In professional programmes can use routine procedures appropriately, but identify situations requiring innovative solutions and draw on relevant theoretical and practical insights in response... (p.17)

These requirements highlight the importance of critical thinking as a foundational skill that underpins students' academic, professional, and social development. By fostering these competencies, higher education institutions play a pivotal role in preparing students to navigate the complexities of contemporary society. Not only that but a Thai workforce which is capable of high-level critical thinking combined with a reasonable wage expectation is a highly attractive prospect for foreign companies and offshore investment. In this manner, a high level of critical thinking would be a net benefit to the Thai economy and to Thai wider society.

Despite prescribing the development of students' critical thinking abilities, the Thailand Qualifications Framework (TQF): Higher Education (HEd) provides a limited elaboration around what the term 'critical thinking' means. How do practitioners know that they understand and are talking about the exact same thing? The vague abstractions stated in the document have underlined the significance of understanding the foundations of critical thinking and identify what constitutes critical thinking skills. Without a coherent explanation and guidance, there seems to be a little by way of teaching methodology for integrating critical thinking within subject content. The vague abstractions stated in the national curriculum have highlighted the importance of understanding the foundations of critical thinking and determining what we mean by it and what constitutes critical thinking skills. Ennis (1987) and Kuhn (1999) assert that a clear definition of critical thinking and its elements are important for the teaching and learning of the subject.

The definitions of critical thinking have been developed and extended by philosophers and thinkers through time. However, there is a coherent theme in the majority of definitions which highlight the significance of evaluating ideas, gathering relevant information and evaluating evidence to support conclusions. For example, Dewey (1910), who contributed to modern critical thinking in the 20th century, accentuated the difference between 'reflective thinking' and 'ordinary thinking' and defined 'reflective thinking' as:

Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends. (Dewey, 1910, p. 6)

Extending Dewey's conception of critical thinking into an educational context, it is vital for teachers to provide students with opportunities to practice reflection and critical thinking skills. More importantly, this underlines the need for teachers who are trained in teaching

critical thinking skills and who are confident in incorporating the critical thinking process into their teaching.

Theorists and researchers' views of critical thinking and its elements vary across different fields. Critical thinking in philosophy centres around logical thinking skills whilst in psychology, research focuses more on cognitive skills and disposition. In education, the concept of critical thinking is mainly related to creating the conditions for learning critical thinking. The ability of thinking critically in an educational context is not merely to recall information and basic concepts, explain ideas or concepts, or apply information in new situations. Rather, it involves the upper three hierarchical levels of Bloom's Taxonomy, including examining information and drawing connections amongst ideas and justifying decisions or conclusions. Although critical thinking is not exactly equivalent to the higher order thinking skills (Ennis, 1987), several scholars (e.g. Ennis, 1987; Fisher, 1990; Halpern, 2007; Ruggiero, 1988) have advocated that critical thinking is a part of or at least a requirement of higher-order thinking which is, in itself, generally defined as a cognitive ability and a practical ability. According to a consensus arrived by the Delphi panel of 46 experts of critical thinking, critical thinking is viewed as a collective skills which composed of the six core elements, including, interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 2015).

Prior to designing an appropriate instructional tool to develop students' critical thinking, an understanding of the levels of a student's critical thinking abilities seems to be a fundamental requirement. To illustrate, a data set of the current levels of students' critical thinking abilities is essential for indicating certain areas of weakness and strength in critical thinking, and the information will be of benefit for further designing and developing instructional tools to enhance their critical thinking abilities. Regarding assessing the levels of students' critical thinking, Ennis (1993, p. 180) points out that "if we are to know where to focus our instruction, we must start with where they are in specific aspect of critical thinking (*sic*)".

There is relatively little research in literature addressing a measurement or an assessment of the cognitive ability levels using a widely recognised test that can indicate the critical thinking capabilities of Thai students. Many research studies on critical thinking skills within the Thai context tend to focus on investigating critical thinking abilities of Thai students using various instructional instruments. For example, Ploysangwal (2018) employed an analytical and critical reading test to investigate Thai students' critical thinking abilities. There are a number of published literature studies investigating the impact of instructional tools such as debate in second or foreign language learning and reporting on its contribution to the significant development of students' critical thinking abilities (e.g. Aclan & Aziz, 2015; Iman, 2017; Jost, 2018). In the Thai context, Liu (2017) investigated the effects of Thai high school students' critical thinking skills after implementing the debate instruction in a flipped learning environment. In another aspect, many research studies examined Thai students' perceptions on the significance of critical thinking skills and dispositions (e.g. Padmarudram Sasidharan Thampi et al., 2024). Due to the shortage of literature about critical thinking skill assessment within the Thai context; therefore, I endeavoured to find out whether there are valid and effective critical thinking tests on the market that Thai students can reliably be assessed on.

Globally recognised tests which measure critical thinking abilities with a multiple-choice format include Watson-Glaser Critical Thinking Appraisal (WGCTA), Cornell Critical Thinking Test (CCTT), and California Critical Thinking Skills Test (CCTST), for instance. Amongst these three tools, CCTT Level Z has been used to assess seven subskills,

including induction, deduction, creditability, assumptions, semantics, definition, and prediction in planning experiment. Different from CCTT, CCTST evaluates seven subskills in analysis, interpretation, inference, evaluation, explanation, induction, deduction, and numeracy (quantitative reasoning). The WGCTA test tends to mostly align with six core elements, as identified by the Delphi panel. According to Watson and Glaser's indicators, the test consists of reading texts and multiple-choice questions that measure and assess a range of critical thinking abilities, including drawing inferences, recognising assumptions, deducing, interpretating, and evaluating arguments.

The Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) has been recognized as one of the reputable global tests for its validity and reliability[1] in diagnosing the critical thinking abilities of high school and university students. A study by Clifford et al. (2004) informed the reliability of the WGCTA test. In the study, 101 undergraduate students took the WGCTA test, and the coefficient alpha was 0.76. Likewise, in Hassan and Madhum's (2007) study, 273 Lebanese students took the WGCTA test and the reliability coefficient of the test scores was 0.74. Barnett and Francis' research (2012) studied the validity and reliability of the WGCTA and found that the reliability coefficient of the test scores of 147 students was 0.71. The above findings confirmed that the WGCTA test is an appropriate tool for an assessment of individuals' critical thinking abilities on a large scale. Despite widespread use of the WGCTA test, there seems to be very little or almost no empirical research exists investigating the critical thinking abilities of Thai students using this WGCTA test.

Effective assessment with the WGCTA test will enable educators to gauge Thai undergraduate students' current levels of critical thinking skills and refine pedagogy and instruction to further support their development. Although it is considered that the WGCTA test is an effective critical thinking assessment tool, directly employing the WGCTA test for the Thai context can have significant challenges. That is to say, the use of the WGCTA test, for which the texts and the question items are written in English, raises issues around the difficulties Thai test takers might encounter when reading the test in L2. Some published evidence shows a correlation between English language proficiency and a critical thinking ability. For instance, Rashid and Hashim's (2008) research study inferred a positive correlation between critical thinking abilities and the language proficiencies of 280 Malaysian undergraduate students. The overall results indicated that the students who were proficient in English language appeared to exhibit more of a critical thinking ability. It can be assumed that an individual's weak English language proficiency can result in her limited ability in developing and exercising critical thinking. The findings lend support to the view that a critical thinking test in the Thai language is required for Thai test takers. That is, to avoid any second language barrier which can affect the comprehension of test takers and critical thinking test results, this researcher postulated that a critical thinking test should be written in the first language of the test taker. Although the WGCTA test is available in a wide range of languages, including UK English, US English, French, Dutch, Spanish and Portuguese, there is no original design that has been translated into the Thai language. Accordingly, the use of the WGCTA test to measure Thai students' critical thinking abilities optimally would require a Thai language translation. As one of the researchers is a Thai university lecturer teaching English as a foreign language, the authors saw significant benefits in working closely with the publisher of a WGCTA test to produce the test in a Thai language edition. However, our proposal for the collaborative project is unlikely to be of interest to the publisher. There was a limited possibility to purchase the WGCTA test online

and to create a different derivative language version owing to copyright reasons and this then limited the possibility to use the test as a reference for forming the questions.

There was the need to create a reliable, modified, open-access critical thinking test for Thai undergraduates. This research study was intended to develop a critical thinking test in Thai language modelled on the WGCTA test format being used for diagnosing the current levels of critical thinking abilities of Thai undergraduates. Such a test would hopefully make a critical thinking assessment of Thai undergraduates possible using a modification of the open-access critical thinking exercises which helps prepare test takers for the WGCTA test. This research article aims to report the results of testing a prototype of a critical thinking test in its Thai translation in the pilot stage of the study.

METHODOLOGY

Design-Based Research (DBR) holds significant potential for empowering researchers to design and develop instruments aimed at measuring the critical thinking abilities of Thai students. As a practical research methodology, DBR combines theoretical foundations with real-world applications by utilizing an iterative approach. This iterative process allows researchers to refine and enhance their tools and strategies continuously, ensuring they are both effective and contextually relevant. According to Anderson and Shattuck (2012), DBR is particularly valuable because it generates the necessary knowledge to guide and improve educational research and practice.

DBR has been applied in a wide range of educational settings, demonstrating its versatility and adaptability. These include traditional onsite classrooms, online learning environments, and even workplace training settings, as noted by Zheng (2015). The methodology is particularly characterized by its focus on designing, implementing, and evaluating interventions through multiple cycles of testing and refinement. Each iteration provides opportunities to learn from practical applications, leading to progressively improved outcomes.

Another key aspect of DBR is its emphasis on generating actionable and practical design principles. These principles are aimed not only at enhancing educational practices in specific, localized contexts but also at providing insights that can be adapted to similar contexts elsewhere. By bridging the gap between theory and practice, DBR contributes to the advancement of both educational research and the real-world practices it seeks to improve. This dual focus makes DBR a robust and flexible approach for addressing complex challenges in education.

Research design

The DBR model was a major factor in formulating this research project. According to the model, there are four stages, including (1) analysis of practical issues, (2) design of tools/solutions, (3) testing and refinement of tools/solutions, and (4) reflection. The first stage of the DBR model allowed researchers to identify a practical issue within the Thai university context. While the development of students' critical thinking has been emphasised, understanding which critical thinking subskills students possess and lack is essential for effective pedagogical design and instruction. There appeared to be a lack of research on the current levels of critical thinking abilities among Thai undergraduate

students. This finding highlighted the need to develop a critical thinking test in the Thai language to assess Thai undergraduates' critical thinking abilities in the second stage. Following the principles of a DBR model, the initial testing of the instrument was conducted as a pilot study. This stage was critical for evaluating the instrument's performance and identifying areas for improvement before proceeding to subsequent iterations. The participants in the first pilot test were provided with opportunities to provide feedback on the test and the accompanying manual of the test. The feedback of the pilot participants and the analysis of the test reliability were used to inform a refinement of the tools for its second iteration. A systematic trial of the revised tools would then be carried out in its third iteration. The final application of the DBR model involves a reflection on the outcomes of the third iteration in order to further improve the critical thinking test and the instructions in order to generate the design principles for the tool development which would then inform the Thai context and other similar contexts.

Participants

To refine the instrument before full-scale testing, the first iteration included a pilot study involving 52 volunteers. The cohort consisted of Thai undergraduate students currently enrolled in the English as a Foreign Language (EFL) programme at a university located in northern Thailand. To facilitate participant selection, the researchers employed a convenience sampling method. The recruitment process was further simplified by relying on the students' willingness to volunteer for the research project, ensuring ethical participation. This approach was chosen due to the accessibility of the target group and the researchers' ability to engage with these students efficiently. Instead of accessing the entire target population within a limited timeframe, this method allowed the researchers to first gain an understanding of the target audience's feedback before conducting full-scale research. This smaller, manageable group allowed the researchers to identify potential limitations, weaknesses, or areas for improvement in the test-instrument. Insights gained from this pilot phase would then serve as a foundation for revising and enhancing the use of the tool and its results, ensuring effectiveness and reliability in subsequent testing iterations.

Instruments

The following sections illustrate the overview and the rationale of the research instruments which took the form of a critical thinking test with an accompany explanatory manual.

Critical Thinking Test

The design of the paper-and-pencil critical thinking test in this research was inspired by the adaptation and modification of online Watson-Glaser Critical Thinking Appraisal (WGCTA) practice tests. These practice tests served as a foundation for constructing a tool that would effectively measure a broad spectrum of critical thinking abilities. The WGCTA practice test demonstrated its capacity to assess various essential skills, including drawing inferences, recognizing assumptions, making deductions, interpreting information, and evaluating arguments. Structured into five subtests, the WGCTA comprises a total of 40

question items. Each subtest begins with a text, which is followed by a series of questions that require specific critical thinking sub-skills to answer accurately.

For the purpose of this research, the researchers selected statements from the WGCTA subtests, focusing on those with clear and comprehensible content as well as common knowledge which would be accessible to the target participants. To ensure linguistic and cultural relevance, the selected statements were translated into Thai, maintaining their original meaning and intent. The first stage of this study involved assessing the initial version of the critical thinking test to support its further improvement. Rather than undergoing a lengthy back-translation process with a panel of experts, I was able to facilitate the Thai translation by drawing on my expertise and experience in the field, given my dual role as a lecturer in the Department of English and a researcher. Adhering to the structure of the WGCTA, the critical thinking test developed for this study was also divided into five sections or subtests, namely: 1) evaluation of arguments, 2) identifying assumptions, 3) making deductions, 4) interpreting information, and 5) drawing inferences.

Each subtest contains four texts accompanied by eight multiple-choice questions, resulting in a total of 40 question items across all sections. This design ensures comprehensive coverage of critical thinking skills while being accessible to the participants. The test structure was carefully designed to align with the original WGCTA format, enabling a systematic assessment of students' critical thinking abilities in a structured and organized manner.

The WGCTA traditionally allows 30 minutes to complete 40 questions. For this edition of the critical thinking test, the time limit was extended to 40 minutes, providing participants with an average of one minute per question. This adjustment aimed to mitigate issues associated with speeded tests, particularly since some participants may not have been familiar with these kinds of critical thinking tests. A tight time limit can negatively affect reliability if participants are unable to complete the test, leading to results that do not accurately reflect their true abilities. By offering additional time, participants were given the opportunity to carefully consider each question item.

Accompanying Explanatory Manual

To address the possibility that some participants were unfamiliar with the Watson-Glaser Critical Thinking Appraisal (WGCTA) test or had never taken any similar critical thinking assessments before, it was crucial to develop a comprehensive accompanying manual. This manual was carefully adapted from the original English version of the WGCTA test and translated into Thai to ensure that it was accessible and comprehensible for all participants.

The primary purpose of the manual was to provide clear instructions, detailed explanations, and illustrative examples for each type of critical thinking skill assessed in the WGCTA test. The manual began with an overview of the objective and structure of the test, laying a solid foundation for understanding. It was systematically organized into five subtests or sections, each focusing on a distinct aspect of critical thinking: 1) evaluation of arguments, 2) assumptions, 3) deductions, 4) interpreting information, and 5) inferences. Each section commenced with a concise explanation of the critical thinking subskill it covered, ensuring participants clearly understood the focus of that particular part of the test. This explanation included definitions and practical illustrations of each subskill in question.

Since the WGCTA test uses a multiple-choice format, the manual provided guidance on how to approach the questions, offering insights into the reasoning or conditions that should guide the selection of a particular answer choice. To further enhance understanding,

each subtest included one example text followed by two illustrative question items. For each question, the manual provided not only the correct answer but also a detailed explanation of why that answer was correct. This approach aimed to deepen the participants' comprehension of the reasoning process behind each critical thinking task.

The structure and format of the manual being applied consistently across all sections would then ensure a seamless and coherent learning experience. By systematically incorporating clear explanations, examples, and answer rationales, the accompanying manual was designed to prepare participants effectively, boosting their familiarity with the test format and their confidence in tackling the questions.

Data collection and analysis

This research study was reviewed and approved by the Chiang Mai University Research Ethics Committee (CMUREC 65/222). To ensure ethical compliance, all students willing to participate in the pilot study were provided with an information sheet outlining the details of the research study. Informed consent was obtained from each participant prior to their involvement. Participants were also explicitly informed of their right to withdraw from the study at any time without any repercussions, and their responses were kept strictly anonymous to maintain confidentiality.

The pilot study involved 52 volunteer participants and was designed to identify potential limitations or weaknesses in the instrument for further refinement. At the beginning of the session, each participant received a copy of the accompanying manual, which served as an introduction to the critical thinking subskills and provided examples to familiarize participants with the test format. A 20-minute period was allocated for participants to review the manual and absorb its concepts and examples. During this time, they were allowed to ask for clarification if they had any doubts. Participants kept the manual and were permitted to refer to it while taking the critical thinking test.

Following this preparation, the participants were given the critical thinking test and instructed to complete it within 40 minutes. The time frame was set to simulate realistic testing conditions while maintaining a balance between thoroughness and practicality. After completing the test, participants were encouraged to provide feedback and comments on the manual, offering insights which hopefully would be invaluable for improving its clarity, content, and overall usability for subsequent iterations.

The data collected from the pilot study, including the test scores, were analyzed using SPSS. Descriptive statistical methods, such as calculating the mean, standard deviation (SD), and percentage, were employed to summarize the data. Additionally, the reliability of the test was examined using the alpha coefficient, which assessed the internal consistency of the instrument. The standard error of measurement was also calculated to evaluate the precision and dependability of the test results.

This systematic approach to the pilot study not only facilitated the identification of potential issues in the instrument but also laid the groundwork for refining and improving its design for future iterations of the research.

RESULTS

Characteristic of the study participants

A total of 52 undergraduate students majoring in English were recruited to participate in this study. The demographic composition of the participants revealed that the majority were female, accounting for 63.46% of the total group. This gender distribution highlights a higher representation of female students among the participants, which aligns with common trends in humanities-related fields such as English studies.

In terms of academic standing, 51.92% of the participants were second-year students, making up the largest proportion of the group. This suggests that the study primarily targeted students who were in the early stages of their academic journey, potentially offering insights into their developing critical thinking skills.

When analyzing the participants' academic performance, as measured by their grade point averages (GPAs), the majority of students fell within the top GPA range of 3.51 to 4.00, indicating a high level of academic achievement among the group. This was followed by a notable proportion of participants with GPAs between 3.26 and 3.50, further emphasizing the strong academic performance of most of the study's participants. These GPA distributions provide a useful context for interpreting the findings, as they reflect the participants' academic aptitude, which may influence their performance on the critical thinking test.

Overall, this demographic and academic profile provides a clear overview of the participants, allowing for a better understanding of the population studied and the potential factors that may have impacted the results.

Test reliability

The analysis of the critical thinking test during its first iteration revealed that the instrument exhibited low reliability. Specifically, the overall Cronbach's alpha value for the test was calculated to be 0.171, signifying a very low level of internal consistency among the test items. Cronbach's alpha is a widely used statistical measure for assessing the reliability of a scale, with values ranging from 0.00 (indicating no consistency) to 1.00 (indicating perfect consistency).

A higher Cronbach's alpha value reflects stronger relationships among the items on the test, suggesting that the items measure the same underlying construct in a consistent manner. In research, a Cronbach's alpha value of 0.7 or higher is generally considered acceptable for indicating adequate internal consistency. The observed value of 0.171, however, falls far below this threshold, signaling significant issues with the test's ability to produce stable and reliable results.

This low reliability suggests that the participants' scores on the test were inconsistent, making it difficult to draw meaningful conclusions about their critical thinking abilities based on the test results. It also implies that the relationships between the items on the test were weak, meaning that the items may not have effectively measured the intended critical thinking subskills or may have been interpreted inconsistently by the participants.

Interestingly, the low reliability could also be attributed to the relatively homogeneous abilities of the participants in the specific skills the test aimed to measure. If the participants were nearly equal in their critical thinking abilities, even small variations in their scores could

significantly impact their rankings within the group. This lack of variability among participants may have contributed to the low Cronbach's alpha value.

The findings from this analysis highlight the need for substantial refinement of the test. Adjustments could include revising or eliminating poorly performing items, enhancing the clarity and relevance of the questions, or increasing the number of items to better capture the range of critical thinking skills. These steps would aim to improve the test's internal consistency and its ability to reliably assess critical thinking abilities in future iterations.

Accompanying explanatory manual

The pilot study also sought feedback from the participants on the explanatory manual for improvement. After the test, they could comment on its clarity, structure, and usefulness. This feedback helped identify areas for enhancement to better support understanding of the critical thinking test and its requirements. The manual's initial version had five sections, each covering a critical thinking subskill. Each section included a text, a question, the correct answer, and an explanation to familiarize participants with the test's focus and reasoning process.

The feedback from the participants revealed a few areas for improvement. While some minor revisions were required to correct typographical errors, four participants provided similar comments and suggestions for enhancing the booklet. One of the most frequently mentioned recommendations was the need to include additional examples within each section. They suggested that having more question items, along with corresponding answers and detailed explanations, would help them better understand the critical thinking subskills being tested. This would also give them more opportunities to practice and become more comfortable with the format and reasoning required for the actual test.

The feedback emphasized the importance of providing ample examples to ensure participants are well-prepared and confident in navigating the critical thinking test. Based on these insights, revisions to the question booklet would focus on expanding the number of illustrative examples, ensuring clearer explanations, and addressing any additional concerns raised by the participants. These changes would aim to make the booklet a more effective preparatory tool for future iterations of the study.

Analysis of difficulty index

Table 1 presents the difficulty index scores for the critical thinking test. The mean difficulty index score was 0.62, which indicates that, on average, the test items were relatively easy. This overall ease may have contributed to the low reliability, as tests that are too easy or too difficult are less effective at distinguishing between high- and low-performing test-takers.

When examining the difficulty index of individual test items, it was found that 17 out of 40 items (42.5%) were classified as very easy. A test with a significant proportion of very easy items risks failing to measure the full range of participants' abilities, as most test-takers are likely to answer these questions correctly and this then reduces the variability in scores and impacts the test's ability to reliably assess critical thinking skills. Similarly, it was noted that nine out of 40 items (22.5%) were classified as very difficult, making them equally problematic. Test items that are too difficult can frustrate participants and may result in random guessing, which also decreases the reliability of the test. Interestingly, the

proportions of very easy and very difficult items were nearly equal, with each category comprising a significant portion of the test. However, items at an appropriate level of difficulty—those that effectively balance challenge and accessibility—were scarce. Only two out of 40 items (5%) were found to meet the ideal difficulty threshold, highlighting a significant imbalance in the test's design.

Based on these findings, a substantial revision of the test items is necessary. Items classified as very easy and very difficult should be carefully evaluated and, where possible, excluded or revised. While eliminating such items outright may reduce the number of test items, doing so is essential for improving the test's overall reliability and ability to differentiate between varying levels of ability. At the same time, it is acceptable to retain items classified as relatively difficult and very difficult, provided they are well-designed and contribute meaningfully to the assessment of critical thinking skills. These items can add to the test's discriminatory power, especially when balanced with a greater number of items at an appropriate difficulty level.

Table 1. Index of Critical Thinking Test Difficulty

Items	No. of Correct Answers (R)	Difficulty (P)	Meanings	Suggestions
Q1	52	1.00	Very easy	Need to be excluded
Q2	51	0.98	Very easy	Need to be excluded
Q3	34	0.65	Relatively easy	Fair to be included
Q4	9	0.17	Very difficult	Need to be excluded
Q5	42	0.81	Very easy	Need to be excluded
Q6	23	0.44	Appropriate	Keep it
Q7	47	0.90	Very easy	Need to be excluded
Q8	47	0.90	Very easy	Need to be excluded
Q9	22	0.42	Appropriate	Keep it
Q10	46	0.88	Very easy	Need to be excluded
Q11	4	0.08	Very difficult	Need to be excluded
Q12	4	0.08	Very difficult	Need to be excluded
Q13	46	0.88	Very easy	Need to be excluded
Q14	3	0.06	Very difficult	Need to be excluded
Q15	15	0.29	Relatively difficult	Fair & can be included
Q16	48	0.92	Very easy	Need to be excluded
Q17	50	0.96	Very easy	Need to be excluded
Q18	36	0.69	Relatively easy	Fair & can be included
Q19	19	0.37	Relatively difficult	Fair & can be included
Q20	47	0.90	Very easy	Need to be excluded
Q21	37	0.71	Relatively easy	Fair & can be included
Q22	8	0.15	Very difficult	Need to be excluded
Q23	41	0.79	Relatively easy	Fair & can be included
Q24	49	0.94	Very easy	Need to be excluded
Q25	45	0.87	Very easy	Need to be excluded
Q26	47	0.90	Very easy	Need to be excluded
Q27	44	0.85	Very easy	Need to be excluded
Q28	51	0.98	Very easy	Need to be excluded
Q29	36	0.69	Relatively easy	Fair & can be included
Q30	16	0.31	Relatively difficult	Fair & can be included

Q31	46	0.88	Very easy	Need to be excluded
Q32	33	0.63	Relatively easy	Fair & can be included
Q33	33	0.63	Relatively easy	Fair & can be included
Q34	34	0.65	Relatively easy	Fair & can be included
Q35	4	0.08	Very difficult	Need to be excluded
Q36	9	0.17	Very difficult	Need to be excluded
Q37	50	0.96	Very easy	Need to be excluded
Q38	10	0.19	Very difficult	Need to be excluded
Q39	6	0.12	Very difficult	Need to be excluded
Q40	38	0.73	Relatively easy	Fair & can be included

The current test includes a disproportionate number of very easy and very difficult items, which hinders its effectiveness as a reliable measure of critical thinking. Future iterations should focus on developing and incorporating a greater number of moderately challenging items to create a more balanced and effective instrument. By doing so, the test will better differentiate between participants' abilities, ultimately enhancing the reliability and validity of its results.

Overall, most test items requiring advanced critical thinking skills are classified as very easy or relatively easy. Additionally, the majority of these items demonstrate limited ability to differentiate between high-performing and low-performing test takers. Therefore, it is essential to review and revise the critical thinking test items before administering them in its second iteration.

Analysis of discrimination index

In addition, the discrimination index of the critical thinking test was designed to evaluate the ability of test items to distinguish between high-scoring and low scoring test takers. As shown in Table 2, the discrimination index for 34 out of 40 items was very low, indicating that the test overall lacks discriminatory efficiency. This suggests that most test items were ineffective in differentiating between the abilities of those who potentially could have performed well and those who potentially would have performed poorly. Notably, six test items had values close to zero, meaning that most participants achieved similar scores on those items.

DISCUSSION

Limitations and recommendation

One limitation of this study is the absence of the classical back-translation method, which is typically used to assess the accuracy and fidelity of a translation to the source text. This limitation arises from the approach taken in directly translating the critical thinking disposition instrument for Thai undergraduates from English into Thai. Given my dual role as a lecturer in the Department of English and a researcher, the translation into Thai was made feasible by leveraging my knowledge and experience in this field. This research study aimed to evaluate the critical thinking test in its first iteration to facilitate further refinement. However, due to time constraints and the lengthy process involved in conducting back-

translation with a panel of experts, it was deemed more practical to postpone the implementation of the back-translation method. The researcher suggests that this method be applied after establishing test-retest reliability and finalizing the critical thinking test. Consequently, it is recommended to conduct further studies aimed at comparing and validating the English version of the WGCTA with its Thai translation. Additionally, future research should focus on a comparative analysis of the original English version and the Thai-translated version to ensure the accuracy and preservation of content.

Table 2: Discrimination Index of Critical Thinking Test

Items	No. of Successes (R)	No. of Successes in High Group (H)	No. of Successes in Low Group (L)	Discrimination (r)	Meanings
Q1	52	26	26	0	Very low, revision needed
Q2	51	26	25	0.02	Very low, revision needed
Q3	34	18	16	0.04	Very low, revision needed
Q4	9	4	5	-0.02	Very low, revision needed
Q5	42	22	20	0.04	Very low, revision needed
Q6	23	13	10	0.06	Very low, revision needed
Q7	47	26	21	0.10	Relatively low, revision needed
Q8	47	24	23	0.02	Very low, revision needed
Q9	22	16	6	0.19	Relatively low, revision needed
Q10	46	23	23	0	Very low, revision needed
Q11	4	4	0	0.08	Very low, revision needed
Q12	4	1	3	-0.04	Very low, revision needed
Q13	46	25	21	0.08	Very low, revision needed
Q14	3	1	2	-0.02	Very low, revision needed
Q15	15	10	5	0.10	Relatively low, revision needed
Q16	48	26	22	0.08	Very low, revision needed
Q17	50	25	25	0	Very low, revision needed
Q18	36	20	16	0.08	Very low, revision needed
Q19	19	13	6	0.06	Very low, revision needed
Q20	47	26	21	0.10	Relatively low, revision needed
Q21	37	21	16	0.10	Relatively low, revision needed
Q22	8	6	2	0.08	Very low, revision needed
Q23	41	21	20	0.02	Very low, revision needed
Q24	49	26	23	0.06	Very low, revision needed
Q25	45	22	23	-0.02	Very low, revision needed
Q26	47	25	22	0.06	Very low, revision needed
Q27	44	24	20	0.08	Very low, revision needed
Q28	51	26	25	0.02	Very low, revision needed
Q29	36	18	18	0	Very low, revision needed
Q30	16	10	6	0.08	Very low, revision needed
Q31	46	23	23	0	Very low, revision needed

Q32	33	18	15	0.06	Very low, revision needed
Q33	33	17	16	0.02	Very low, revision needed
Q34	34	20	14	0.12	Relatively low, revision needed
Q35	4	1	3	-0.04	Very low, revision needed
Q36	9	8	1	0.13	Relatively low, revision needed
Q37	50	25	25	0	Very low, revision needed
Q38	10	7	3	0.08	Very low, revision needed
Q39	6	4	2	0.04	Very low, revision needed
Q40	38	21	17	0.08	Very low, revision needed

Another limitation of this study is the lack of an investigation into the validity of the critical thinking test. Specifically, the draft version of the test should be assessed for both content validity and face validity. For content validity, a panel of experts could evaluate the relevance of each item in relation to the range of critical thinking abilities measured by the WGCTA, using a four-point Likert scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant). The draft version of the critical thinking test should also be assessed for face validity. Test takers could rate the clarity and comprehensibility of the test items using a four-point Likert scale (1 = not clear, 2 = somewhat clear, 3 = quite clear, 4 = very clear).

Another notable limitation of this research is the lack of diversity among participants. All participants were selected from the same undergraduate program at a single university, resulting in a homogeneous group that exhibited limited variability in critical thinking abilities. This lack of participant diversity may have constrained the generalizability of the findings, as the results reflect only the critical thinking dispositions and abilities of a narrowly defined population. To address this limitation, future studies should aim to include a broader and more diverse pool of participants. Specifically, participants should be recruited from various undergraduate programs across different disciplines and, ideally, from multiple universities. Such diversity would not only enhance the reliability and validity assessments of the critical thinking test by incorporating a wider range of perspectives and abilities but would also offer valuable insights into the critical thinking dispositions of students from different academic and cultural backgrounds. This approach could help identify variations in critical thinking skills across disciplines, potentially informing the development of more targeted and effective educational interventions. Furthermore, ensuring a diverse sample could improve the overall applicability and robustness of the test, enabling it to serve as a more universally relevant instrument for evaluating critical thinking abilities.

CONCLUSION

This Thai-translated edition of the critical thinking test is not suitable for further use in assessing the critical thinking abilities of Thai undergraduate students. During its initial iteration, the analysis revealed that the translated instrument demonstrated a low reliability which limited its effectiveness as a measurement tool. Despite this drawback, the unexpected findings from the study offer some significant insights that contribute to the development of effective strategies for creating a robust critical thinking test tailored to the needs of Thai educators and practitioners.

The research study also provides valuable implications for the application of translation methodologies and cultural adaptation processes in test development. It underlines the importance of careful modifications to the critical thinking test through direct translation to align it with the cultural and linguistic context of the target audience. While it is recognized that no critical thinking test can be entirely free from flaws, and that such tests may inherently contain certain errors, the overarching aim of the researcher, as the test developer, is to refine and enhance this test. The ultimate objective is to develop a reliable critical thinking assessment tool with a minimal margin of error, ensuring its applicability and relevance in educational settings.

NOTES

1. Validity of a test refers to the extent to which a test truly measures the knowledge or abilities it is intended to measure. Test reliability is defined as the degree to which a test produces consistent results over repeated administration.

DISCLOSURE STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

Aclan, E. M., & Aziz, N. H. A. (2015). Exploring parliamentary debate as a pedagogical tool to develop English communication skills in EFL/ESL classrooms. *International Journal of Applied Linguistics & English Literature ISSN*, *4*(2), 2200–3592. https://doi.org/10.7575/aiac.ijalel.v.4n.2p.1

Aiken, L. R. (2003). *Psychological testing and assessment* (11th ed.). Allyn and Bacon.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, *32*(2), 201–211. https://doi.org/10.1080/01443410.2011.638619

Clifford, J. S., Boufal, M. M., & Kurtz, J. E. (2004). Personality traits and critical thinking skills in college students: Empirical tests of a two-factor theory. *Assessment*, *11*(2), 169–176. https://doi.org/10.1177/1073191104263250

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment* (8th ed.). McGraw-Hill.

Dewey, J. (1910). *How we think*. D.C. Heath & Co.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). W. H. Freeman and Company.

Ennis, R. H. (1993). Critical thinking assessment. *Thoeory in Practice*, *32*(3), 179–186. http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=a0a11a17-a57e-4d61-bcac-47fa68b41bd6%40sessionmgr4008

Facione, P. A. (2015). Critical thinking : What it is and why it counts. *Insight Assessment*, *ISBN 13: 978-1-891557-07-1.*, 1–28. https://doi.org/ISBN 13: 978-1-891557-07-1

Fisher, R. (1990). *Teaching children to think*. Basil Blackwell Ltd.

Halpern, D. F. (2007). The nature and nurture of critical thinking. In R. J. Sternberg, H. L. Roediger III, & D. F. Halpern (Eds.), *Critical thinking in psychology* (Issue 1, pp. 1–14). Cambridge University Press. https://doi.org/10.1017/CBO9780511804632.002

Hassan, K. EL, & Madhum, G. (2007). Validating the Watson Glaser Critical Thinking Appraisal. *Higher Education*, *54*, 361–383. https://doi.org/10.1007/s10734-006-9002-z

Iman, J. N. (2017). Debate instruction in EFL classroom: Impacts on the critical thinking and speaking skill. *International Journal of Instruction*, *10*(4), 87–108. https://doi.org/10.12973/iji.2017.1046a

Jost, N. (2018). Theoretical support for the use of debate in the Japanese EFL classroom. *Encounters*, *6*, 29–50. https://dokkyo.repo.nii.ac.jp/?action=repository_action_common_download&item_id=1442 &item_no=1&attribute_id=22&file_no=1

Kuhn, D. (1999). A developmental model of critical thinking. *Source: Educational Researcher*, *28*(2), 16–25. http://www.jstor.org/stable/1177186?seq=1&cid=pdf-reference#references_tab_contents%0D

Liu, I. W. (2017). *The effects of debate instruction through a flipped learning environment on speaking ability and critical thinking skills of Thai high school students* [Doctoral dissertation, Chulalongkorn University]. Chula Digital Collections. https://digital.car.chula.ac.th/chulaetd/700/

Office of the Education Council. (2017). *National qualifications framework (Thailand NQF)* (Revised ed). Office of the Education Council, Ministry of Education.

Office of the Higher Education Commission. (2006). *National qualifications framework for higher education in Thailand: Implementation handbook*. http://www.mua.go.th/users/tqf-hed/news/FilesNews/FilesNews8/NQF-HEd.pdf

Padmarudram Sasidharan Thampi, K. P., Ponathong, C., & Yongsorn, C. (2024). Critical thinking skills and dispositions: Perceptions of higher education students in Thailand. *Journal of Education and Innovation, 26*(3),153–167. https://so06.tci-thaijo.org/index.php/edujournal_nu/article/view/268218

Ploysangwal, W. (2018). An assessment of critical thinking skills of Thai undergraduate students in private Thai universities in Bangkok through an analytical and critical reading test. *University of Thai Chamber of Commerce Journal Humanities and Social Sciences*, *38*, 75–91. https://so06.tci-thaijo.org/index.php/utccjournalhs/article/view/158002

Rashid, R. A., & Hashim, R. A. (2008). The relationship between critical thinking and language proficiency of Malaysian undergraduates. *EDU-COM International Conference*, *November*, 19–21. https://ro.ecu.edu.au/ceducom/36/

Ruggiero, V. R. (1988). *Teaching thinking across the curriculum*. Harper & Row.

Watson, G. B., & Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal manual, forms A and B*. Psychological Cooperation.

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education*, *2*(4), 399–420. https://doi.org/10.1007/s40692-015-0036-z

BIODATA

Tanyapon Phongphio, PhD: A lecturer at Chiang Mai University, Thailand. Her research focuses on developing critical thinking skills among Thai undergraduates. She holds master's degrees in TEFL from Thammasat University and in English Studies from University of Nottingham, along with a PhD in Education from University of Bristol.

Beatriz Gallo Cordoba, PhD (https://orcid.org/0000-0002-7335-4534): A Senior Research Fellow at the Centre for International Research on Education Systems (CIRES) at the Mitchell Institute, Victoria University, Australia. She has over 12 years of experience in the analysis of administrative and survey data using statistical techniques to study multiple problems in educational research.

How to cite this article:

Phongphio, T., & Gallo Cordoba, B. (2025). Developing instrument to measure critical thinking abilities of Thai undergraduate students. *Journal of Language and Culture, 44*(1), 1-17.