TPM Vol. 32, No. 4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

DEVELOPMENT OF A COMPETENCY ASSESSMENT CENTER MODEL FOR MEASURING AND EVALUATING THAI LANGUAGE TEACHERS' CLASSROOM PERFORMANCE

GITIMA SOMVONGSA[1]
JATUPHUM KETCHATTURAT[1]
DAWRUWAN THAWINKARN[1]
[1]KHON KAEN UNIVERSITY

KEOW NGANG TANG[2]
[2]NILAI UNIVERSITY

Abstract: This study aimed to develop, design, and evaluate a Measurement and Evaluation Competency Assessment Center Model to assess Thai Language teachers' competencies in real-world. A total of 45 Thai Language teachers and five experts with more than 10 years experience as participants. A research and development design was employed for creating a Quality Assessment Toolkit testing, and refining them through experts' validation and field trials. Four instruments were document analysis checklist, content validity form, scoring rubric, and satisfaction questionnaire. The process encompassed needs analysis, model design, expert validation, and pilot testing. The results indicated four components of teachers' competencies: planning for measurement and assessment, creation and quality of measurement tools, classroom measurement and evaluation, and reporting on learning outcomes and applying assessment results. Finally, expert evaluation confirmed the model has very high levels of suitability, feasibility, possibility, utility, accuracy, and accountability. Hence, it is clear, relevant, and practical to use.
Keywords: Competency assessment model;

INTODUCTION:

Teacher quality is consistently identified as one of the strongest school-based predictors of student learning outcomes. Recent work on teacher competency evaluation frames it as a systematic process for accessing teachers' knowledge, skills, and professional dispositions, with the aim of improving instruction and ensuring accountability (Simonson et al., 2021). In other words, competence in this sense is not limited to content knowledge, but also includes planning and organizing instruction, using appropriate assessment strategies, managing classroom, and engaging in professional collaboration (Raha, 2024).

In specific language education such as Thai Language, effective classroom performance involves facilitating meaningful communication, cultivating high-order thinking, and providing feedback that supports language development (Kotsunthon, 2025). Past studies of teaching competence highlight that high-performing teachers demonstrate clear instructional goals, flexible use of teaching strategies, and a strong capacity to monitor and respond to students in real time. These dimensions are particularly salient for Thai Language teachers, who must simultaneously develop students' linguistic accuracy, reading and writing skills, and appreciation of Thai culture and identity (Apridayani et al., 2025).

Thailand's Basic Education Core Curriculum B.E. 2551 (A.D. 2008) defines key student competencies and desirable characteristics, which in turn shape expectations for teacher curriculum. The curriculum emphasizes holistic development and specifies core competencies that students should achieve, including communication, thinking, problem-solving, and technological skills. In order to support these aims, teachers are expected to design learning activities that foster active participation, critical thinking, and application of knowledge to real-life situations. Following this line of reasoning, a Competency Assessment Center Model is a powerful mechanism for improving teacher quality in Thailand and it is significantly important to investigate (Kotsunthon, 2025).

The Competency Assessment Center Model is importance to reflect a competency-based education approach, where Thai Language teachers can assess their students not only on content knowledge but also on instructional planning, learner engagement, assessment literacy, and use of the Thai Language in varied contexts. This helps to ensure Thai Language teachers can apply their knowledge effectively in real classroom settings (Adarsh et al., 2021). Adarsh et al. (2021) highlighted that a Competency Assessment Center Model is useful in ensuring a systematic, fair, and meaningful evaluation of Thai Language teachers, hence it is expected to contribute to broader goals of educational excellence and equity. Following this line of reasoning, a Competency Assessment Center Model is vital for Thai Language teachers, in particular, in the context of improving

education quality, ensuring teacher effectiveness, and aligning with national education reforms (Jameson et al., 2022).

Competency-based education refers to a learner-centered approach that emphasizes measurable outcomes, where progression is based on demonstrated performance of desired competencies, rather than seat time or traditional grading (Sultan et al., 2025). In other words, competency-based education is an educational approach that focuses on the mastery of specific skills, knowledge, and behaviours, so-called competencies rather than time-based progression such as grade levels. This means that competency-based education needs to make sure that learners can only advance once they have demonstrated that they have achieved clearly those defined learning outcomes (Sultan et al., 2025). According to Do and Nguyen (2021), teachers should be equiped with knowledge and skills to select, adapt, and design classroom assessment tools following curriculum-based competencies to help learners in developing their 21st-century skills.

In Thai Language education, measurement and evaluation are not merely tools to provide learners' scoring, but they are integral processes that shape teaching, enhance learning quality, and support continuous improvement in both learners and teachers (Pholying, 2025). As a result, educational measurement and evaluation are used to ensure the Thai Language learning outcomes are accurately monitored, systematically improved, and meaningfully achieved. Therefore, assessment and evaluation serve as crucial pillars that underpin instructional effectiveness and learner learning outcomes in the evolving landscape of education. The dichotomy between traditional assessment methods and emerging competency-based approaches highlights significant shifts in pedagogical practices in Thailand (Kotsunthon, 2025).

Thailand's education policies progressively emphasize quality teaching and accountability (Kolb & Kolb, 2022). Therefore, a Competency Assessment Center Model can align with initiatives from the Office of the Basic Education Commission and Teacher Council of Thailand to standardize teaching quality and professional standards (Raha & Wongsapan, 2024). According to Zhang et al. (2023), assessment centers can help the Thai Language teachers to meet national goals by guaranteeing them in possessing the required knowledge, skills, and attitudes, thus improving their learners' learning outcomes. In addition, a Competency Assessment Center Model can provide detailed feedback on individual teachers' strengths and areas for improvement, which can inform them regarding teacher's learning plans, targeted training, and career progression decisions (Dobinson & Dockrell, 2021). Following this line of reasoning, Thai Language teachers will be benefited from structured opportunities in terms of self-reflection, continuous learning, and professional growth, hence enhances these Thai Language teachers' professional development (Chukwunemerem, 2023). Furthermore, traditional teacher competencies evaluations often rely on subjective observations or outdated tools so a Competency Assessment Center Model uses multiple evidence-based methods, such as simulations, interviews, classroom performance tasks, can assess Thai Language teachers' competencies, making the evaluation more accurate, more transparent, and less biased (Asyari et al., 2021). This is crucial for educational administration in improving the validity and fairness of teacher evaluation for them to make a fair and appropriate promotion decisions, in particular (Quaigrain & Arhin, 2017).

There are several key problems that related to competency assessment for Thai Language teachers, in particular, in the context of Thailand's education system. Current situation is lacking of Standardized Competency Framework to assess Thai Language teachers' competencies. In other words, there is no universally accepted or consistently applied framework that clearly defines the core competencies for Thai Language teachers (Chantarat & Chookhampaeng, 2023). Some schools may have their own interpretation, leading to inconsistent assessment criteria (Sukkamart et al., 2025). Many competencies' assessments rely on rote memorization or written examinations, not measuring practical teaching capabilities, classroom management skills, and inspiration of student engagement (Child & Ellis, 2021), which is considered outdated assessment methods. Moreover, performance-based assessments, such as lesson demonstrations or learner feedback, are underused (Papanthymou & Darra, 2022). Current assessments often emphasize linguistic knowledge over instructional competence such as ability to design lesson plans, differentiate instruction, and use of digital tools, so-called insufficient focus pedagogical skills (Smirmova, 2023). This creates a research gap between what teachers know and how effectively they can teach is questionable.

Based on the above problems and research gaps, the researchers sought to achieve the following research questions: (i) To develop a set of competency assessment tools focused on the measurement and evaluation skills required for Thai Language teachers, according to professional standards and educational best practices; (ii) To design and implement a Measurement and Evaluation Competency Assessment Center Model as a practical framework for assessing Thai Language teachers' competencies in real-world and simulated teaching contexts, and (iii) To evaluate the quality and effectiveness of the Measurement and Evaluation Competency Assessment Center Model in terms of its validity, reliability, feasibility, and user satisfaction among Thai Language teachers, assessors, and educational administrators.

METHOD

Research Design

The researchers employed a Research and Development design which is ideal for creating educational tools, testing them, and refining them through expert validation and field trials (Thakur, 2021). The research process consisted of four phases. The first phase was research on need analysis whereby the researcher's reviewed literature on measurement and evaluation competencies for Thai Language teachers. Then, the researchers conducted document analysis to identify key competencies for their components and

indicators and followed by analyzing gaps in current assessment practices. In second phase, the researchers developed a draft set of Quality Assessment Toolkit comprised of (i) a general test; (ii) a test of writing skills; (iii) measurement and evaluation design skills test; (iv) feedback skills test; (v) attitude measurement form, and (vi) manual of using the Measurement and Evaluation Competency Assessment Center Model. The researchers sought content validation from experts using Content Validity Index. This was followed by the researchers revised based on the five experts' feedback.

The third phase was to design a Measurement and Evaluation Competency Assessment Center Model which included design the structure, process, stations, assessment methods and tools as well as prepare manuals, assessor guidelines, and scoring criteria. After that, the researchers conducted a pilot test with a sample group of 45 Thai Language tecahers and trained assessors for consistency in scoring. In this trial implementation, the researchers collected data on implementation, performance, and teacher feedback. In the final phase, an evaluation of effectiveness and quality of Measurement and Evaluation Competency Assessment Center Model was conducted to assess reliability, validity, practicality, and satisfaction. Based on the evaluation results, the researchers revised the tools and assessment center structure based on trial results and proposed guidelines for scaling the use of the Measurement and Evaluation Competency Assessment Center Model.

Population and Samples

The population consisted of 262 Thai Language teachers who are working in opportunity expansion schools under the jurisdiction of the Chaiyaphum Primary Educational Service Area 2, Office of Basic Education Commission, academic year 2023. The researchers employed a purposive sampling to select 45 samples to participate in this study (Ahmed, 2024). Other than this, a total of five experts with not less than 10 years experience in various specializations were selected to examine the appropriateness and feasibility on the components and indicators of Measurement and Evaluation Competency Assessment Center Model (Andrade, 2020). They are: (i) an expert from Department of Research and Development, Mahasarakham University; (ii) a director from Thai Language Institute, Office of Academic Affairs and Educational Standards, Office of the Basic Education Commission; (iii) a lecturer from Department of Business Management, Faculty of Management, Ubon Ratchathani University; (iv) an expert from Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University, and (v) an expert from Department of Educational Leadership and Management, Faculty of Education, Khon Kaen University.

Research Instruments

A total of four research instruments were used to collect data at every phase of the study. Based on the research design and objectives, the researchers required to have a variety of research instruments tailored for different phases of the research, starting from tool development to implementation and evaluation of the Measurement and Evaluation Competency Assessment Center Model. In the need analysis phase, the researchers used document analysis checklist to review existing competency standards and curriculum frameworks to identify key competencies and understand current gaps.

In the tool development and validation phase, the researchers used Content Validity Form to validate and refine the Measurement and Evaluation Competency Assessment Center Model. The five subject matter experts were required to assess each item's relevance, clarity, and appropriateness. During designing the Measurement and Evaluation Competency Assessment Center Model and piloting the implementation, the researchers utilized performance tasks or scenarios to assess teacher competencies in real and simulated settings. The realistic teaching and assessment-related activities such as designing a test, analyzing student work, and interpreting data were used as research instruments. Besides, a scoring rubric with clear performance levels and descriptors for each competency was employed. In the evaluation phase, a satisfaction questionnaire was used to assess the quality of the tools and the Measurement and Evaluation Competency Assessment Center Model for Thai Language teachers and their assessors respectively, to evaluate clarity, fairness, usefulness, and engagement (Awu, 2021).

Data Collection and Data Analysis

In the first phase, a document analysis was conducted using a systematic review of relevant literature to identify the essential components and indicators of the Measurement and Evaluation Competency Assessment Center Model (Glenn, 2009). This phase served as the foundation for constructing the assessment tools and determining the competency domains to be evaluated (Glenn, 2009). Sources of documents were selected from peer-reviewed academic journals, books, policy papers, international standards set by Organization for Economic Cooperation and Development (OECD), United Nations Educational Scientific and Cultural Organization (UNESCO), and Thailand National Curriculum Frameworks that related to teacher education, educational measurement, as well as professional competency assessment. Documents were included if they addressed competency frameworks or assessment in higher education or teacher education, provided conceptual or operational definitions of competencies, included measurable indicators or assessment practices, and have been published within the past 10 years.

The selected documents were analyzed using content analysis to extract and categorize information on competency components and performance indicators. The document analysis followed these steps: (i) The documents were read thoroughly to gain an overall understanding as initial reading; (ii) This was followed by coding whereby relevant text segments were coded based on recurring themes related to competencies (for example, writing feedback, evaluation design, or attitudes); (iii) Codes were grouped into broader categories representing core competency domains, so-called thematic grouping; (iv) Next was indicator extraction such as specific indicators, namely observable behaviors or skills were extracted and refined under each

domain, and (vi) Validation on the identified components and indicators were cross-checked across multiple sources to ensure consistency and comprehensiveness (Glenn, 2009). These components and indicators were then used to guide the development of assessment instruments and the overall structure of the Quality Competency Assessment Center toolkit.

In the second phase of this study, a multi-stage data collection was carried out in order to prepare and examine the Quality Competency Assessment Center toolkit. The toolkit consisted of six elements: (i) a 35-questions of general test; (ii) a Writing Skills Examination; (iii) a measurement and evaluation design skills test; (iv) a feedback skills test; (v) an attitude measurement form, and (vi) a manual for using the Measurement and Evaluation Competency Assessment Center Model. Data were collected to support the development, validation, and refinement of each element of the Quality Assessment Toolkit, as described below.

The researchers defined clearly the aim of the 35-questions of general test was to measure Thai Language tecahers' competencies. This was followed by compiling a list of questions to be used in the general test, in order to make sure that each question aligns with the defined aim. The five experts assessed each question for its alignment with the aim, using a rating scale from -1 to +1 (1 = consistent, 0 = somewhat consistent, and -1 = not consistent). The researchers determined the passing criteria as the questions that meet the evaluation criteria should have an Item Objective Congruence (IOC) value of at least 0.5. Those questions with an IOC value lower than 0.5 may need revision or replacement to better align with the research objectives (Tuner & Carlson, 2003). The index of IOC was used to assess the alignment between test items and intended objectives of a 35-questions of general test for the Quality Assessment Toolkit. It is a quantitative method in test development to determine how well a test item measures a specific objective, as judged by the five subject matter experts. The IOC analysis helps to identify items that are not well-aligned with the objectives, potentially impacting the validity of the general test (Johari et al., 2011).

Besides IOC value, the difficulty of index was used to evaluate the quality of the 35-questions of the general test. The difficulty index would tell the researchers how 'easy' or 'hard' the general test question is the questions, and it is particularly useful in test validation and quality assurance (Johari et al., 2011). In short, the difficulty index also called p-value in classical test theory and it is the proportion of learners who answered the item correctly. The difficulty index ranges from 0.0 to 1.0. The higher value interprets that the question is easy while the lower value interprets as the question is difficult. Table 1 elucidates the interpretation of difficulty index (Johari et al., 2011).

TABLE 1 Interpretation of Difficulty Index

Difficulty index	Interpretation	Action
0.80 to 1.00	Very easy	Consider revising and add challenge
0.60 to 0.79	Easy	Usually acceptable
0.40 to 0.59	Moderate	Ideal range for most tests
0.20 to 0.39	Difficult	Review for clarity or appropriateness
0.00 to 0.19	Very difficult	Likely flawed or too advanced

.

The discrimination index is another important tool used in test analysis. It assists the researchers to evaluate how well the general test question distinguishes between high-performing and low-performing students. It is often used alongside the difficulty index to improve the quality of assessment (Quaigrain & Arhin, 2017). Following this line of reasoning, discrimination index measures how effective a question differentiates between learners who performed well overall (the upper group) and those who did poorly (the lower group). Therefore, a high discrimination index means the item is good at telling apart strong and weak learners (Quaigrain & Arhin, 2017). Table 2 demonstrates the interpretation guide for discrimination index (Quaigrain & Arhin, 2017).

TABLE 2 Interpretation Guide for Discrimination Index

Discrimination index	Interpretation	Action
0.40 and above	Excellent	Keep item
0.30 to 0.39	Good	Acceptable
0.20 to 0.29	Fair	May need revision
0.00 to 0.19	Poor	Review and revise
Negative	Very poor	Item may be flawed (reverse key, ambiguity, etc.)

.

The Kuder-Richardson Formula 20 (KR-20) was employed by the researchers to assess the reliability (internal consistency) of the 35-questions of general test with binary items, for example, questions scored as right or wrong (1 or 0). It is especially useful in educational testing when evaluating multiple-choice or true/false questions (Quaigrain & Arhin, 2017). KR-20 tells the researchers how consistently the test measures the same concept or skill. A high KR-20 score means the test items are reliably measuring the same construct. Table 3 shows the interpretation guide for KR-20 (Quaigrain & Arhin, 2017).

TABLE 3 Interpretation Guide of KR-20

KR-20 value		Interpretation
≥0.90		Excellent (very reliable)
0.80 to 0.89		Good
0.70 to 0.79		Acceptable
0.60 to 0.69		Questionable
<0.60		Poor – test may need revision

To evaluate appropriateness of examination content of the Quality Assessment Toolkit, the researchers used mean score ranges. Learners' average performance on each examination item or section was used to determine whether the content is appropriate or not for the target level (Quaigrain & Arhin, 2017). The researchers used predefined mean score ranges to interpret appropriateness, as elucidated in Table 4.

TABLE 4 Interpretation of Appropriateness for Examination Content

Mean score range	Interpretation	Content appropriateness
0.00 to 0.30	Very difficult	Too hard; likely inappropriate
0.31 to 0.50	Difficult	Challenging; may need revision
0.51 to 0.70	Moderate	Acceptable appropriate
0.71 to 0.90	Easy	Appropriate but could be more rigorous
0.91 to 1.00	Very easy	Too easy; may not assess learning properly

.

Cronbach's Alpha (α) is a measure of internal consistency used to evaluate how well the items in a test or questionnaire measure the same underlying construct, for example knowledge, attitude, skill (Quaigrain & Arhin, 2017). The researchers based on the General Interpretation Scale, as illustrated in Table 5 to interpret the Cronbach's Alpha value for reliability testing.

TABLE 5 Cronbach's Alpha Interpretation

Cronbach's alpha	Interpretation
≥0.90	Excellent (Very high internal consistency)
0.80 to 0.89	Good
0.70 to 0.79	Acceptable
0.60 to 0.69	Questionable (Needs improvement)
0.50 to 0.59	Poor
<0.50	Unacceptable

.

The researcher used a mean score range based on Likert-scale-based feedback from the 45 Thai Language teachers to evaluate the effectiveness of a prototype Measurement and Evaluation Competency Assessment Center Model in terms of suitability and feasibility. Below are the scoring criteria with mean score ranges assuming a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) (Jebb et al., 2021).

TABLE 6 Scoring Criteria from Mean Score Range (Effectiveness Evaluation)

Mean score range	Interpretation	Effectiveness level
4.50 to 5.00	Very high: Strongly suitable/feasible	Excellent effecriveness
3.50 to 4.49	High: Generally suitable/feasible	Good effectiveness
2.50 to 3.49	Somewhat suitable/feasible	Moderate effectiveness
1.50 to 2.49	Low: Limited suitability/feasibility	Low effectiveness
1.00 to 1.49	Very low: Not suitable/feasible at all	Poor effectiveness

RESULTS AND DISCUSSION

Results of the Document Analysis in First Phase

The results from document analysis checklist in the need analysis phase found out that there are four components and 19 indicators derived from a systematic review of existing existing competency standards, teacher assessment framework, and curriculum guidelines. These components collectively describe the essential domains required for effective classroom measurement and assessment practices among Thai Language teachers. The results confirm the current literature and policy documents emphasize a comprehensive, multi-dimensional approach to assessment literacy, moving beyond basic test construction toward a broader cycle of planning, implementation, interpretation, and utilization of assessment information. Table 7 displays the initial results based on document analysis.

TABLE 7 Identification of Components and Indicators of Measurement and Evaluation Competency Assessment Center Model

Components	Indicators
Component 1: Planning for classroom measurement and assessment	Indicator 1.1: Defines the objectives of classroom measurement and assessment.
	Indicator 1.2: Analyzes student learning outcomes according to the standards and indicators in the curriculum.
	Indicator 1.3: Develops an assessment plan that is consistent with student learning outcomes.
	Indicator 1.4: Use a variety of methods, techniques, and tools to measure and evaluate in the classroom.
	Indicator 1.5: Sets the criteria for classroom evaluation.
Component 2: Creation and quality of classroom measurement and evaluation tools	Indicator 2.1: Diagram of the structure and characteristics of classroom measurement and evaluation tools.
	Indicator 2.2: Design tasks / work items for use in assessing students' classroom performance.
	Indicator 2.3: Create tools for measuring and evaluating in the classroom according to desirable characteristics.
	Indicator 2.4: Examine and analyze the quality of classroom measurement and evaluation tools.
Component 3: Classroom measurement and evaluation operation	Indicator 3.1: Communicates details of measurement and assessment to students so that they understand the objectives, tools, methods, criteria, tasks, and scores according to the assessment plan.
	Indicator 3.2: Assessment and analysis of individual student.
	Indicator 3.3: Assesses progress during learning to focus on monitoring students' development.
	Indicator 3.4: Assesses post-learning achievement to focus on students' learning outcomes at the end of semester.
	Indicator 3.5: Collect and analyze data, and summarize students' classroom evaluation results.
	Indicator 3.6: Determines learning outcomes according to the specified assessment criteria.
Component 4: Reporting on learning outcomes and applying assessment results	Indicator 4.1: Prepare a report of information and applying assessment results.
	Indicator 4.2: Assesses students' progress during their learning to monitor their development.
	Indicator 4.3: Assesses post-learning achievement to focus on students' learning outcomes at the end of the semester.
	Indicator 4.4: Collect and analyze data, and summarize the results of student classroom assessments.

Component 1: Planning for Classroom Measurement and Assessment

The first component identified relates to Thai Language teachers' capacity to plan assessment activities purposefully and systematically. The presence of relevant indicators in the reviewed documents suggests that effective assessment begins with clear alignment between learning outcomes, instructional activities, and assessment strategies. This result resonates with international frameworks that argues planning is the foundation for valid and reliable assessment practices. The five indicators found such as establishing objectives, choosing suitable assessment, and aligning assessment with curriculum standards, reflect a strong emphasis on assessment design coherence. This alignment is crucial in competency-based and outcome-based education models, ensuring that assessment practices genuinely measure intended competencies.

Component 2: Creation and Quality of Classroom Measurement and Evaluation Tools

The second component focuses on teachers' ability to design, construct, and evaluate the quality of measurement tools. Indicators related to test blueprinting, item construction, rubric development, and ensuring validity and reliability were prominent across the reviewed documents. This result indicates that high-quality assessment instruments depend heavily on Thai Language teachers' technical skills in measurement. It also highlights a common challenge reported in previous studies that many teachers struggle with developing instruments that meet psychometric standards. The identification of this component suggests that the existing competency standards recognize the necessity for strong technical assessment skills, particularly in the context of modern competency-based curricula that require diverse forms of evidence.

Component 3: Classroom Measurement and Evaluation Operation

The third component concerns the actual implementation of assessments in the classroom. The indicators reflect Thai Language teachers' ability to administer assessments ethically, manage assessment conditions, collect data accurately, score student work consistently, and use appropriate tools and

technologies to support assessment activities. The emphasis on operational competencies suggests a shift from traditional testing toward more dynamic, classroom-based measurement practices. This aligns with broader educational trends where teachers are expected to employ a range of authentic and formative assessment methods, monitor student progress continuously, and ensure fairness throughout the assessment process. The results further indicate that operational skills are important for transforming well-designed assessment tools into meaningful classroom evidence.

Component 4: Reporting on Learning Outcomes and Applying Assessment Results

The final component emphasizes the interpretation, communication, and application of assessment information. Indicators include analyzing student data, reporting results clearly to stakeholders, providing constructive feedback, and using assessment evidence to improve instruction. This component underscores the transformative purpose of assessment, not merely generating grades, but using test results to enhance learning and teaching. The prominence of such indicators in existing standards affirms assessment as part of a continuous improvement cycle. The results also reveal the increasing expectation for teachers to be data-literate and to use assessment outcomes for differentiated instruction and learning intervention planning.

Results of the Quality Assessment Toolkit in Second Phase

The Quality Assessment Toolkit consists of 35 questions of a general test, an examination of writing skills, measurement and evaluation design skills test, feedback skills test, attitude measurement form, and a manual for using the Measurement and Evaluation Competency Assessment Center Model to assess Thai Language teachers' competencies in the real-world contexts. Table 8 illustrates the results of the 35 questions of general test in terms of its IOC values, difficulty index, discrimination index, and KR-20 score. The IOC value was found in the range of 0.80 to 1.00 which passed the quality criteria at least 0.5. Therefore, the IOC value showed that all the 35 questions of general test are aligning with the research objectives; hence, the results enhance the reliability of the obtained data (Quaigrain & Arhin, 2017). The results of IOC imply that the 35 questions can increase confidence, reduce measurement errors, and improve the general test quality.

Furthermore, the difficulty index is used to evaluate the quality of the 35 questions. The difficulty index with the values ranges from 0.40 to 0.73 showed that the questions are easy to moderate (refer to Table 1). Thus, the results imply that most of the questions fall in the ideal range and usually are acceptable. Next, the results of discrimination index ranges from 0.20 to 0.53 indicated that some questions may need revision, some are acceptable, and some can be kept (refer to Table 2). Finally, the KR-20 score showed a great reliability result (refer to Table 3).

The evaluation of the Quality Assessment Toolkit, consisting of 35 general test items and several subcomponents such as writing skills test, measurement and evaluation design skills test, feedback skills test, attitude measurement form, and user manual, demonstrates strong evidence of measurement quality and alignment with the intended assessment purposes. The results in Table 8 highlight four key psychometric indicators, namely IOC, difficulty index, discrimination index, and KR-20 reliability.

Firstly, the IOC values ranging from 0.80 to 1.00 exceed the minimum accpetable threshold of 0.50, indicating that expert reviewers consistently judged all 35 items as highly congruent with the research objectives and competency constructs. This suggests strong content validity, affirming that the test items adequately reflect the targeted domains of measurement and evaluation competencies for Thai Language teachers. As highlighted by Quaigrain and Arhin (2017), high IOC values contribute to greater confidence in the instrument, reduce measurement errors, and enhance the overall validity of the data obtained.

Secondly, the difficulty indices between 0.40 to 0.73 indicate that the majority of items fall within the easy-to-moderate difficulty range. Items within this interval are generally considered ideal, as they neither overwhelm lower-performing respondents nor fail to challenge higher-performing ones. This implies that the general test is appropriately calibrated for the target population and can effectively differentiate respondents across ability levels without causing test fatigue or ceiling/floor effects.

Thirdly, the discrimination index values ranging from 0.20 to 0.53 reveal that while many items function adequately in distinguishing high- and low-performing respondents, some may require refinemnet. Items with discrimination values at or below 0.20 typically warrant revision or removel due to their limited ability to distinguish between levels of competence. Conversely, items closer to 0.50 demonstrate stronger discriminative power and can be retained. These results suggest that although the majority of the items are acceptable, a targeted item review would further enhance the test's discriminative strength.

Finally, the KR-20 reliability coefficient indicates a high level of internal consistency among the 35 items, demonstrating that the general test reliably measures a single competency construct. A high KR-20 value suggests that the items are well-correlated and consistently capture the underlying measurement and evaluation competencies expected of Thai Language teachers. This strong reliability further confirms the robustness of the Assessment Toolkit as a standardized measure suitable for real-world competency evaluation.

TABLE 8 Quality Results of 35 Questions of General Test

Quality testing	Value	Criteria of quality
IOC	0.80 to 1.00	≥0.5 (pass)
Difficulty index	0.40 to 0.73 (easy to moderate)	Ideal range and usually acceptable (pass)

Discrimination index	0.20 to 0.53	Keep item (pass)
KR-20	0.86	Good (pass)

The researchers employed inter-rater agreement index as a statistical measure used to determine the consistency or agreement between different raters assessing the same performance – in this case, writing skills in the examination (Quaigrain & Arhin, 2017). Cohen's Kappa was used to the two raters utilizing agreement beyond chance, ranged from – 1 (complete disagreement to 1 (perfect agreement) and 0 will be chance agreement, according to Jebb et al.'s (2021) interpretation as shown in Table 9.

TABLE 9 Inter-rater Agreement Index: Cohen's Kappa (K)

Inter-rater agreement index	Interpretation
<0	Poor
0.01 to 0.20	Slight
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 1.00	Almost perfect

.

The results of inter-rater agreement index based on Cohen's Kappa (K)'s interpretation indicated as 0.88. This means that the inter-rater agreement index was found almost perfect on the Examination Writing Skills Test. Besides, the overall suitability was found at a high level too (mean score = 4.46; mean range = 0.89; standard deviation = 0.13). Content appropriateness was found appropriate but could be more rigorous (refer to Table 4) (Landis & Koch, 1977).

On the other hand, quality testing results for design skills test for measurement and evaluation in the classroom showed that the appropriateness was at high level (mean score = 4.52; mean range = 0.90; standard deviation = 0.16). Therefore, content appropriateness was found appropriate but could be more rigorous (refer to Table 4), with an inter-rater agreement index based on Cohen's Kappa (K) of 0.84. Referring to Table 9, it means the inter-rater agreement index was found almost perfect on the design skills test for measurement and evaluation in the classroom.

Next, quality testing results of a feedback skills test indicated a high level of appropriateness (mean score = 4.52; mean range = 0.90; standard deviation = 0.08). This implies that a feedback skills test was found appropriate but could be more rigorous in terms of its content, with an inter-rater agreement index based on Cohen's Kappa (K) of 0.86. This value implies that it was found almost perfect on the feedback skills test.

The fifth element of Quality Assessment Toolkit is the attitude scale. The IOC value was found at 0.80 to 1.00. Since the IOC values exceeded 0.5, passing the evaluation criteria and identified the attitude scale was better aligned with the research objectives (Turner & Carlson, 2003). Moreover, the reliability value of Cronbach Alpha was 0.76, was found it was at acceptable internal consistency (refer to Table 5). The final element of Quality Assessment Toolkit is the manual for using the Measurement and Evaluation Competency Assessment Center Model. Based on a 5-point rating scale results showed that the appropriateness was at high level (mean score = 4.53; mean range = 0.90; standard deviation = 0.22). Therefore, content appropriateness was found appropriate but could be more rigorous (refer to Table 4), with an inter-rater agreement index based on Cohen's Kappa (K) of 0.85. Referring to Table 9, it means the inter-rater agreement index was found almost perfect on the manual for using Measurement and Evaluation Competency Assessment Center Model.

Results of the Implementation a Prototype of the Measurement and Evaluation Competency Assessment Center Model in Real School Settings

In final phase, the researchers intended to evaluate the effectiveness of the Measurement and Evaluation Competency Assessment Center Model in terms of its suitability and feasibility. This phase was mainly to confirm whether the Measurement and Evaluation Competency Assessment Center Model is able to provide accurate, fair, and context-sensitive tools for measuring and evaluating teaching competencies in Thai Language classroom (Quaigrain & Arhin, 2017). A total of 45 Thai Language teachers from opportunity expansion schools under the jurisdiction of the Chaiyaphum Primary Educational Service Area 2, Office of Basic Education Commission participated in the implementation a prototype of Measurement and Evaluation Competency Assessment Center Model in real school settings. The results of the implementation the prototype 45 Thai Language teachers' perspective is demonstrated in Table 10.

TABLE 10 Results of 45 Thai Language Teachers on the Quality of Measurement and Evaluation Competency Assessment Center Model

Teacher competencies	Suitability			Possibility		
	Mean	*SD*	Interpretation	Mean	*SD*	Interpretation

Planning for classroom measurement and assessment	4.60	0.55	Very high	4.80	0.45	Very high
Creation and quality of classroom measurement and evaluation tools	4.80	0.45	Very high	4.40	0.55	High
Classroom measurement and evaluation operation	4.60	0.55	Very high	4.60	0.55	Very high
Reporting on learning outcomes and applying assessment results	4.80	0.45	Very high	4.60	0.55	Very high
Feedback summary	4.60	0.55	Very high	4.40	0.55	Very high
Average	4.68	0.11	Very high	4.56	0.09	Very high

Besides, the 45 Thai Language teachers' results on the quality of Measurement and Evaluation Competency Assessment Center Model, the researchers also looked for the five experts' evaluation too. The purpose of the experts' evaluation was to assess the quality of the developed Measurement and Evaluation Competency Assessment Center Model in terms of its utility, feasibility, propriety, accuracy, and accountability for measuring competencies among the target group, e.g., Thai Language teachers. The results are presented in Table 11 below in a clear, structured, and evidence-based using mean scores.

TABLE 11 Results of the Measurement and Evaluation Competency Assessment Center Model According to the Five Experts' Evaluation

Assessment standards	Mean	*SD*	Interpretation
Utility standards	4.35	0.33	Very high
Feasibility standards	4.60	0.45	Very high
Propriety standards	4.65	0.38	Very high
Accuracy standards	4.65	0.22	Very high
Accountability standards	4.70	0.21	Very high

The overall results from the five experts' evaluation indicated that the Measurement and Evaluation Competency Assessment Center Model is of high quality and demonstrates strong validity and suitability for assessing competencies in the targeted educational context. Minor revisions based on experts' feedback have enhanced the model's clarity and practicality.

CONCLUSION

This study aimed to develop a Measurement and Evaluation Competency Assessment Center Model specifically designed to measure and evaluate the classroom performance of Thai Language teachers. Through a systematic process that included needs analysis, model design, expert validation, and pilot testing, the study successfully produced a comprehensive and contextually relevant assessment model grounded in competency-based education principles. The final Measurement and Evaluation Competency Assessment Center Model comprise multiple components that integrate both performance-based tasks and structured evaluation tools aligned with key teaching competencies such as planning for classroom measurement and assessment, creation and quality of classroom measurement and evaluation tools, classroom measurement and evaluation operation, and reporting on learning outcomes and applying assessment results. In addition, expert evaluations confirmed the Measurement and Evaluation Competency Assessment Center Model's very high levels of suitability, feasibility, possibility, utility, accuracy, and accountability, with quantitative results indicating

strong agreement among assessors regarding the clarity, relevance, and practicality of the Measurement and Evaluation Competency Assessment Center Model.

The results affirm that the developed Measurement and Evaluation Competency Assessment Center Model is a reliable and effective tool for systematically assessing Thai Language teachers' real-world teaching performance. It provides a structured framework for formative and summative evaluation, supports professional development, and aligns with national standards for teacher quality. In conclusion, the Measurement and Evaluation Competency Assessment Center Model offer significant potential for enhancing the quality of teaching and learning in Thai Language classrooms. Its adoption by education institutions from primary up to tertiary education, hence, teacher training programs can contribute to more evidence-based, transparent, and competency-focused teacher evaluation practices. Future research should explore its scalability, inter-rater reliability, and long-term impact on teaching effectiveness across different regions and educational levels.

REFERENCES

1. Adarsh, K., Krishnamurthi, R., Bhatia, S., Kaushik, K., Ahuja, N. J., Nayyar, A., & Masud, M. (2021). Blended learning tools and practices: A comprehensive analysis. *IEEE Access, 9*, 85151-85197.
2. https://doi.org/10.1109/access.2021.3085844.
3. Ahmed, S. K. (2024). How to choose a sampling technique and determining sample size for research: A simplified guide for researchers. *Oral Oncology Report, 12*, 100662.
4. https://doi.org/10.1016/j.oor.2024.100662
5. Andrade, C. (2020). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine, 43*, 86-88. https://doi.org/10.11/025371720977000
6. Apridayani, A., Sakkanayok, K., & Sukarsih. (2025). Teacher origin and its influence on Thai students' English proficiency and cross-cultural competence. *International Journal of Multicultural Education, 27*(2), 69-87.
7. Asyari, L., Nuriyanti, R., Gunawan, D., & Adiredja, R. K. (2021). *The influence of Experiential Learning Model on primary school student's creative thinking skills*. https://doi.org/10.20961/shes.v4i1.48568
8. Awu, I. O. (2021). Research instruments: A questionnaire and an interview guide used to investigate the implementation of higher education objectives and the attainment of Cameroon's vision 2035. *European Journal of Educational Studies, 8*(7), 113-127. https://doi.org/10.46827/ejes.v8i7.3808
9. Chantarat, T., & Chookhampaeng, C. (2023). Problems with teaching and learning Thai Language for communication competency for elementary school students. Journal of Educational Issues, 9(1), 36-48. https://doi.org/10.5296/jel.v9i1.20524
10. Child, S., & Ellis, P. (2021). The what, why, and how of assessment: A guide for teachers and school leaders. Sage Publications, Thousand Oaks, California, United States of America. https://www.torrossa.com/en/resources/an/5018839
11. Chukwunemerem, O. P. (2023). Lessons from self-directed learning activities and helping university students think critically. Journal of Education and Learning, 12(2), 79-87.
12. https://doi.org/10,5539/jel.v12n2p79
13. Dobinson, K. L., & Dockrell, J. (2021). Universal strategies for the improvement of expressive language skills in the primary classroom: A systematic review. First Language, 41, 527-554.
14. https://doi.org/10.1177/0142723721989471
15. Glenn, B. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9, 27-40. https://doi.org/10.3316/QRJ0902027
16. Jameson, J., Rumyantseva, N. L., Cai, M., Markowski, M., Essex, R., & McNay, I. (2022). A systematic review and framework for digital leadership research maturity in higher education. Computers and Education Open, 3, 100115. https://doi.org/10.1016/j.caeo.2022.100115
17. Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert Scale development advances: 1995-2019. Frontiers in Psychology, 4(12), 637547. https://doi.org?10.3389/fpsyg.2021.637547
18. Johari, J., Sahari, J., Wahab, D. A., Abdullah, S., Abdullah, S., Omar, M. Z., & Muhamad, W. (2011). Difficulty index of examinations and their relation to the achievement of program outcomes. Procedia-Social and Behavioral Sciences, 18, 71-80.
19. Kolb, A. Y., & Kolb, D. (2022). Experiential learning theory as a guide for experiential educators in higher education. Experiential Learning and Teaching in Higher Education, 1(1), 38.
20. https://doi.org/10.46787/elthe.v1i1.3362
21. Kotsunthon, A. (2025). Assessment and evaluation in Thai classrooms: A comparative study of traditional and competency-based approaches. Insight into Modern Education, 2(2), 13-26.
22. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorial data. Biometrics, 33(1), 159-174.
23. Papanthymou, A., & Darra, M. (2022). The impact of self-assessment with goal setting on academic achievement: Results of a study on primary students in Greece. Journal of Education and Learning, 12(1), 67-90. https://doi.org/10.5539/jel.v12n1p67
24. Pholying, T. (2025). Embracing global Englishes in Thailand: Linguistic diversity and cultural integration. Asian Journal of Arts and Culture, 5(1), 112-130.

25. Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education, 4(1), 2-11.
26. https://doi.org/10.1080/2331186X2017.1301013
27. Raha, T., & Wongsapan, M. (2024). Development of an instructional model to enhance competency in the Thai Language for Grade 6 students. Journal of Education and Learning, 13(2), 119-129.
28. https://doi.org/105539/jel.v13n2p119
29. Simonson, S. R., Earl, B., & Frary, M. (2021). Establishing a framework for assessing teaching effectiveness. College Teaching, 70(3), 134-145. https://doi.org/10.1080/87567555.2021.1909528
30. Smirmova, Y. K. (2023). Application of eye-tracking (DUET) in the study of cooperation between children with atypical development and adults in the learning process. Forum, 1, 98-99.
31. https://doi.org/10.61365/forum.2023.078.
32. Sukkamart, A., Ployduangrat, J., Prongprommarat, J., & Pimdee, P. (2025). Development and validation of a Competency Framework for Thai External Quality Assessment Supervisory Units. Educational Process: International Journal, 18, e2025454. https://doi.org/10.22521/edupij.2025.18.454
33. Sultan, M. A., Miller, E., & Tikkanen, R. S. (2025). Competency-based education and training for community health workers: A scoping review. BMC Health Services Research, 25, 263.
34. https://doi.org/10.1186/s12913-025-12217-7
35. Thakur, H. K. (2021). Research design. Corvette, New Delhi, India.
36. Turner, R. C., & Carlson, L. (2003). Indexes of Item-Objective Congruence for multidimensional items. International Journal of Testing, 3(2), 163-171. https://doi.org/10.1207/S15327574IJT0302-5
37. Zhang, Y., Wu, X., Zhu, C., & Zhou, J. (2023). Classroom quantitative evaluation: A method of both formative and summative evaluation. Sustainability, 15(3), 1-21. https://doi.org/10.3390/su15031783