Development of a Multidimensional Computerized Adaptive Testing Program for Measuring Digital Competency Level in Higher Education

Asanee Tongsilp[1] and Kamonwan Tangdhanakanond[2*]

[1]Research Center, Kasem Bundit University, Thailand
[2*]Faculty of Education, Chulalongkorn University, Thailand

ABSTRACT

Background and Objectives: Thailand has implemented policies to promote and develop digital competencies among students, which require all universities to assess students' digital competencies prior to graduation. As a result, many small universities face resources shortages for conducting such assessments. Therefore, the development of an online testing system serves as a potential solution to this issue. This research aimed to develop a digital competency item bank and validate the effectiveness of a Multidimensional Computerized Adaptive Testing (MCAT) program.

Methodology: This research employed a research and development (R&D) design. Participants consisted of fourth-year undergraduate students enrolled in higher education institutions in Thailand during the academic year 2023–2024, from the fields of Health Sciences, Science and Technology, and Humanities and Social Sciences. The results indicated that the digital competency item bank contained 211 items and 30 indicators. The item bank was developed using a three-parameter Multidimensional Item Response Theory (MIRT) model. The test demonstrated an EAP reliability of 0.865. Digital competency levels were interpreted using a Wright Map. Additionally, the MCAT program was developed using PHP and MySQL.

Results: The quality of 211 items in the item bank was analyzed using three-parameter item response theory. The items demonstrated an average discrimination parameter of 1.44, an average difficulty parameter of 0.24, and an average guessing parameter of 0.13. The item bank was categorized into three modules based on the difficulty parameter. The digital competency model was developed in accordance with the between-items multidimensional item response theory with three parameters. Digital competency level cutoff scores were determined by defining a criterion zone on the Wright map at the mean ability threshold (θ_d) of -0.39. The results indicated that 69.27 percent of students met the digital competency criteria. The highest level of digital competency was observed in Component 1, Information and Digital Media Literacy, followed by Component 3, Adaptability to the Digital Transformation Era,

ARTICLE INFO

Article history:
Received 8 January 2026
Revised 16 October 2026
Accepted 19 October 2026

Keywords:
Digital Competency,
Item Bank,
Multidimensional Item Response Theory (MIRT),
Multidimensional Computer Adaptive Testing (MCAT),

while the lowest level was found in Component 2, Skill for using, developing, and problem-solving with digital tools. In addition, the findings demonstrated that the MCAT program accurately evaluated digital competency with only 22 items, using a variable-length termination criterion of SE(θ) < 0.30 (RMSE = 0.112). This approach provides an effective method to reduce the testing burden.

Discussions: The MCAT program demonstrates the feasibility of implementing a reliable and efficient assessment tool in resource-constrained higher education institutions. At the same time, the study suggests a direction for future development in competency assessment on a web-based platform, which enhances the possibilities for implementing MIRT within an educational organization.

Conclusions: The MCAT program demonstrated strong performance across all assessed dimensions, including content validity, indicators, and program structure. The MCAT program consists of an item bank, initial item selection, subsequent item selection, and a termination criterion. The MCAT approach requires less time and fewer items than traditional tests, with the optimal number of items for accurate ability estimation being 22. Overall, the MCAT program provides an appropriate and practical solution for equitable implementation across large, medium, and small universities.

*Corresponding author
E-mail address: asanee.ton@kbu.ac.th

Introduction

Digital technology plays an increasingly important role in daily life and work. Driving the country forward through technological advancement is essential, particularly in the development of digital citizens. Thailand has implemented national policies to advance the country into the Thailand 4.0 era, guided by the 20-year National Digital Development Policy for the Economy and Society (2018–2037). The goal of the policy is to enhance the country's economic competitiveness by promoting innovation and digital technology, thereby equipping citizens across all sectors —especially youth— with the essential knowledge and skills required to live and work in the digital era. This policy is consistent with the announcement from Thailand's Office of the Higher Education Commission (OHEC), which outlines a strategic plan for fostering digital competence and skills among university students, and mandates that all universities evaluate students' digital knowledge and abilities upon graduation. Universities require sufficient resources for assessment, encompassing standardized tests, examination management, invigilators, and budgets. These requirements pose challenges for smaller universities with limited resources. A survey revealed that many Thai universities, particularly small private universities, continue to rely on traditional testing methods due to limited budgets. Therefore,

the Multidimensional Computer Adaptive Testing (MCAT) system can significantly promote students' equitable participation and developing better learning outcomes.

Over the past decade, Computerized Adaptive Testing (CAT) has become increasingly popular and is regarded as a more effective assessment than traditional paper-and-pencil tests. It is now widely used in large-scale assessments, such as the Test of English as a Foreign Language (TOEFL), Graduate Record Examination (GRE), and Graduate Management Admission Test (GMAT). The CAT uses item selection and scoring based on Item Response Theory (IRT), providing more accurate estimates of the candidate's ability (Brossman & Guille, 2014). CAT is based on the assumption that the test is unidimensional. However, a unidimensional test is sometimes inappropriate for practical assessment. Therefore, the concept of Multidimensional Item Response Theory (MIRT) has been developed. Multidimensional analysis is based on two foundational concepts: uni-dimensionality, which includes a composite approach and a consecutive approach, and multi-dimensionality, which includes between-items and within-items. In the within-items multidimensionality model, each item loads onto all dimensions, while in the between-item multidimensionality model, each item loads onto a specific dimension. (Briggs & Wilson, 2003).

Additionally, conventional methods for MCAT utilize MIRT as a foundation (Reckase, 2009). MCAT accounts for multiple latent traits influencing an individual's responses to items (Frey & Seitz, 2009). Shao et al. (2024) developed multidimensional hybrid computerized adaptive testing (MHCAT), which is a flexible approach that incorporates both modular and adaptively selected items with mensional CAT (MCAT). Results indicated that the variable-length MHCAT had higher estimation accuracy than MCAT. Owing to its multidimensional ability structure, MCAT shows increased complexity in item selection and parameter estimation compared to unidimensional CAT. The MCAT program can assist smaller institutions that lack resources for digital competency assessments of students. It also contributes to curriculum development, learning outcomes, and quality assurance in education related to digital knowledge, skills, and attitudes in higher education.

Objectives of the Study

1. To develop a digital competency item bank using a Multidimensional Item Response Theory (MIRT).
2. To develop and investigate the effectiveness of a Multidimensional Computerized Adaptive Testing (MCAT) program.

Review of Literature

Digital Competency in the 21st Century

Digital competency refers to the ability to integrate knowledge, skills, and attributes to use digital technologies effectively, appropriately, safely, creatively, independently, and ethically. Digital competency is recognized as a vital 21st-century skill, specifically in higher education. Students are expected to engage meaningfully in complex digital environments, with their competency extending beyond basic Information and Communication Technology (ICT) skills, to include critical thinking, information evaluation, digital communication, and ethical use of technology. According to the Digital Competence Framework for Citizens (DigComp 2.0) developed by the European Commission (Vuorikari et al., 2016), digital competence

encompasses five main areas: (1) information and data literacy, (2) communication and collaboration, (3) digital content creation, (4) safety, and (5) problem solving. These competencies serve as a vital skill set for daily life and lifelong learning in the 21st century. A synthesis of research indicates that digital competency comprises various components, with key Components including digital literacy, practical skills, and adaptability (e.g., creation, presentation, teamwork, networking). The digital competency model was developed the digital competency model for undergraduates in the 21st century based on the Higher Education Qualifications Framework 2022, the 21st-century learning skills, the Top 10 Skills identified by the World Economic Forum, and the Digital Competence Framework for Citizens (DigComp 2.0) through research synthesis and focus group discussions with academic experts (Fraumeni & Liu, 2021). The digital competency model consists of three components and 30 indicators: (1) media information and digital literacy, (2) skill for using, developing, and problem-solving with digital tools, and (3) adaptive digital transformation, shown in Table 1.

Table 1

The Digital Competence Model in the 21st Century

Components	Content	Indicators
1. Media information and digital literacy	Computer basics, rights, access to digital media, digital communication, digital safety, understanding digital media, ethics in digital society, digital health, digital commerce, digital law	10
2. Skill for using, developing, and problem-solving with digital tools	Computer usage, Internet usage, word processing programs, spreadsheet programs, presentation programs, computer security, digital production design, write program commands, Installing and maintaining computer peripherals and drivers, solve technical problems, skill adjust in the digital age, digital environment management, use creative technology, Computational thinking	14
3. Adaptive digital transformation	Flexibility and adaptability to technology, social collaboration, self-directed learning, entrepreneurship, technology leadership, follow technological trends	6

Multidimensional Item Response Theory (MIRT) Model

Digital competency is a multidimensional psychological variable that integrates knowledge, skills, and attitudes. Naturally, psychological attributes often exhibit complex behavioral structures. These attributes and their associated content are so interrelated that they cannot be separated and may not reflect a unidimensional structure when assessed through Unidimensional Item Response Theory (UIRT). To address these limitations, the MIRT was developed. MIRT models can measure multiple latent traits that affect an individual's responses to test items (Reckase, 2009). The concept of multidimensionality involves two kinds of relationships among latent traits: (1) between-item multidimensionality, where each question measures a single latent trait, yet several correlated traits are present across different questions; and (2) within-item multidimensionality, where individual questions are influenced by multiple, related latent traits. Consequently, using the MIRT model can reduce the number of questions

required to measure multiple constructs, while maintaining a high level of measurement efficiency in comparison to unidimensional models (Frey & Seitz, 2009).

Multidimensional Computer Adaptive Testing (MCAT) Program

Multidimensional Computer Adaptive Testing (MCAT) integrates Components of MIRT and CAT to enhance the efficiency and accuracy of educational and psychological assessments. MCAT automatically selects items based on an examinee's performance. The MCAT process consists of five steps as follows: The first step involves the development of an item bank. To ensure comprehensive content coverage, the item bank must include a sufficient number of high-quality questions, typically ranging from 100 to 200 items (Weiss & Kingsbury, 1984). The second step is to initiate the test by selecting the first item. Since no prior information about the examinee's ability is available at the beginning of the test, the initial item is chosen based on a moderate difficulty level, with a difficulty parameter (b) close to 0. This approach ensures a balanced starting point, allowing the adaptive process to proceed effectively (Thompson & Weiss, 2011). The third step is ability estimation, which involves estimating the examinee's ability. Ability estimation in MCAT is grounded in item response theory. The method for estimating examinees' latent trait parameters should be selected from the Maximum Likelihood Estimation (MLE) method (Lord & Novick, 2008), the Expected A Posteriori (EAP), and the Maximum A Posteriori (MAP) methods. However, when all responses are either correct or incorrect, the EAP and MAP methods are recommended due to their ability to provide estimates with low standard errors (Hambleton & Swaminathan, 1985) The fourth step, item selection refers to the process of determining the method for selecting items from the item bank to be used in the testing process. The goal of item selection is to minimize the variance of the temporary multidimensional ability vector (Frey & Seitz, 2009). Two primary methods are commonly utilized for this purpose, namely, the Maximum Likelihood (ML) method and the Bayesian Fisher (BF) method. These methods can be used in conjunction with any ability estimation approach and can also be combined with multiple item selection strategies (Reckase, 2009). The final step is the termination criterion. In general, there are two primary rules for terminating a test: (1) the fixed-length rule, where the test ends after a predetermined number of items, and the variable-length rule, where the number of items varies (Reckase, 2009): and (2) the standard error of the ability estimate SE(θ), which sets a threshold for the standard error of ability estimation. SE(θ) is the standard deviation of the distribution of the probability of the approximate value of the true abilities. For Multidimensional Computerized Adaptive Testing (MCAT) research, it is commonly found that the termination criterion is based on a SE(θ) set at 0.3 (Wang & Chang, 2011). Additionally, test lengths of 10, 20, 30, 40, 50, and 60 items have been explored (Finch, 2010). However, the optimal number of items for MCAT remains undetermined. Recent research highlights the reliability of computerized adaptive testing for assessing students, including, measurement precision, reduced test length, and time in testing. As shown in Table 2, recent research on the MCAT focuses on developing highly efficient and accurate testing systems capable of comprehensively measuring complex multidimensional latent traits, ranging from core educational domains to specialized contexts such as diagnosing cognitive attributes and measuring daily behaviors for patients (Li et al., 2025; Shao et al., 2024). Ultimately, the success of MCAT depends on the quality of the item pool and a test design appropriate for the objective of diagnosing multidimensional abilities (Wang & Chang, 2011).

Table 2

Summary of Multidimensional Computer Adaptive Testing

Researcher	Testing Approach	Features	Traits Assessed	Findings
Li et al. (2025)	Multidimensional On-the-Fly Assembled Multistage Testing (OMST-M)	Combines MIRT into MST with on-the-fly test assembly	Cognitive ability domains (e.g., math, language)	Precise measurement, effective item exposure control, and flexible test length and content.
Shao et al. (2024)	Multidimensional Hybrid Computerized Adaptive Testing (MHCAT)	Combines CAT and MST under a MIRT framework	Multidimensional latent traits (e.g., various academic skills)	Better item exposure control than traditional MCAT; high estimation accuracy, especially in variable-length formats
Wang et al. (2022)	FAMCAT (Clinical Application) for hospitalized patients	Multidimensional CAT	Cognition, daily activity, mobility assessments.	Precision in measurement and short test time (average 22 items, 6 minutes)
Piton-Goncalves and Aluísio (2012)	MADEPT System (computer-based multidimensional adaptive testing)	Web-based MCAT system for educational applications	Academic abilities across multiple domains	Accurate measurements, immediate patient assessments, and capable of providing personalized assessments.

Validating the Effectiveness of an MCAT Program

Root mean square error (RMSE) measures the accuracy of ability estimation for examinees. It represents the squared difference between the estimated ability and the true ability, reflecting the variance in the ability estimation process (Babcock & Weiss, 2012).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}\left(\hat{\theta}_{j,\mathrm{MCAT}} - \theta_{j,\mathrm{Paper}}\right)^2}{N}}$$

N is the number of examinees, $\hat{\theta}_{j,\mathrm{MCAT}}$ is the estimated ability level for individual *j* calculated by the MCAT program, and $\theta_{j,\mathrm{Paper}}$ is the criterion estimate (observed full-length paper test score) for individual *j*, serving as the baseline proxy for true ability.

Average bias refers to the accuracy of the estimated ability of the examinees. It is calculated as the average difference between the estimated abilities and their actual abilities. This measure also reflects the direction of the estimation, whether it is overestimated or

underestimated. An average bias equals 0 that the estimated abilities are accurate, indicating no average bias (Babcock & Weiss, 2012).

$$\text{Bias} = \frac{\sum_{j=1}^{N}(\hat{\theta} - \theta)}{N}$$

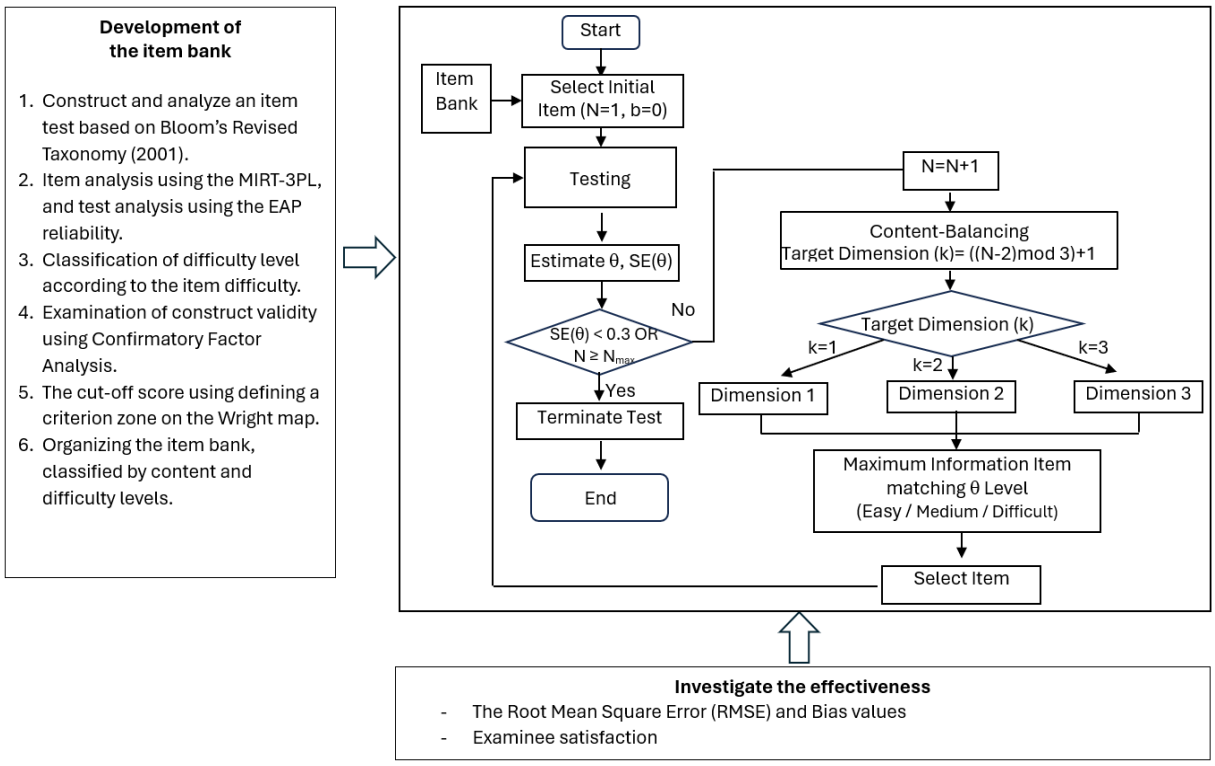
Conceptual Framework

This study involved the development of an MCAT program for testing the digital competence of undergraduate students. The test items were developed by the research team using the frameworks based on Bloom's Revised Taxonomy 2001 (Krathwohl, 2002). The quality of the items was examined using a MIRT-3PL model. This model includes three parameters: (a) discrimination, (b) difficulty, and (c) guessing. The item bank was developed based on content domains and difficulty levels. It was divided into three difficulty levels: (1) easy level ($-3.00 \leq b < -1$), (2) medium level ($-1.00 \leq b \leq 1.00$), and (3) difficult level ($1.00 < b \leq 3.00$), along with content balancing within each difficulty level. The content domains covered were: Component 1: media information and digital literacy, Component 2: Skill for using, developing, and problem-solving with digital tools: Component 3: adaptive digital transformation.

The MCAT program was developed based on the concepts of Reckase (2009) and Thompson and Weiss (2011). In addition, the study aims to compare the accuracy of examinees' ability estimation using two stopping rules: a fixed-length rule of 30 items, and a variable length item with a SE(θ) value < 0.3. The accuracy is assessed using the RMSE and bias values. The conceptual framework is illustrated in Figure 1.

Figure 1

Conceptual Framework of Development of the MCAT Program



Research Methods

Research Design

This research employed a Research and Development (R&D) design, which was divided into two phases: Phase 1 development and quality validation of the item bank and Phase 2 development and effectiveness validation of the Multidimensional Computerized Adaptive Testing (MCAT) program.

Participants

The participants in the study consisted of undergraduate students from universities under the supervision of the Ministry of Higher Education, Science, Research, and Innovation of Thailand.

Participants were divided into two groups. In the first group, a total of 1,256 students were selected using the multistage random sampling method for the development of the item bank. The sample comprised eight out of 155 universities in Thailand, covering the Bangkok metropolitan area, north, south, and central regions. These included one autonomous university, three public universities, and four private universities across three academic disciplines: health sciences, science and technology, and social sciences and humanities. All participants completed the test administration and all were included in the item calibration analysis. For the second group, 112 students were selected in Bangkok selected using the purposive sampling method for the development of the MCAT program from a university that agreed to participate in the online system testing. All participants completed the system trial and were included in the efficacy analysis. Examinees accessed the testing system online.

Instruments

The Digital Competency Test

We developed a digital competency test for a bachelor's degree based on the Higher Education Qualifications Framework 2022, which includes 21st-century learning skills and the Top 10 Skills identified by the World Economic Forum (Fraumeni & Liu, 2021). The test was developed using the MIRT framework, initially consisting of three components, 30 indicators, and 240 multiple-choice items (correct = 1, incorrect = 0). Following the item screening and calibration process based on MIRT criteria, a total of 211 items were retained in the final calibrated item bank (Component 1 = 73 items, Component 2 = 96 items, and Component 3 = 42 items). As presented in Table 3, all 30 indicators remain fully represented, confirming that no indicator was depleted or omitted during the selection process. The content validity was examined by five experts—three in information technology, and two in educational measurement and evaluation. All experts held at least a master's degree and had no less than five years of professional experience in their respective fields. Results revealed that the Index of Item-Objective Congruence (IOC), assessing the alignment of the items with the indicators, the items with learning behaviors, and the questions with answer keys, ranged from 0.67 to 1.00 for all items. Reliability analysis was conducted by calculating the EAP reliability coefficients using the fscores () function in the R program. The overall test revealed an EAP reliability coefficient of 0.865, with coefficients of 0.816 for component 1, 0.827 for component 2, and 0.708 for component 3.

Table 3

Components of the Digital Competency Test

Component 1 Media information and digital literacy			Component 2 Skill for using, developing, and problem-solving with digital tools			Component 3 Adaptive digital transformation		
Indicators	Initial Items	Retained (*N*)	Indicators	Initial Items	Retained (*N*)	Indicators	Initial Items	Retained (*N*)
1. Computer basics	1-8	8	11.Computer usage	81-88	7	25.Flexible	193-200	7
2. Rights	9-16	8	12.Internet usage	89-96	6	26.Social collaboration	201-208	7
3. Access to digital media	17-24	8	13.Word processing program	97-104	7	27. Self-directed learning	209-216	7
4. Digital communication	25-32	8	14.Use spreadsheet program	105-112	6	28.Creation/ entrepreneurship	217-224	7
5. Digital safety	33-40	8	15.Presentation program	113-120	6	29.Technology leadership	225-232	7
6. Understanding digital media	41-48	8	16.Computer security	121-128	5	30. Follow techno-logical trends	233-240	7
7. Ethics in digital society	49-56	6	17. Digital production design	129-136	7			
8. Digital health	57-64	7						
9. Digital commerce	65-72	6	18.Write program	137-144	8			
10.Digital law	73-80	6	19. maintaining computer	145-152	8			
			20.Solve technical problems	153-160	6			
			21. Digital Adaptation	161-168	8			
			22. digital management	169-176	7			
			23. Creative technology	177-184	7			
			24.Computational thinking	185-192	8			
Total		73			96			42

The MCAT System

The MCAT program was developed to measure digital competency at the undergraduate level. It was constructed using the PHP programming language and a MySQL database. The program manages data linkage within the databases and was developed based on the principles of the Software Development Life Cycle (SDLC) (Elliott, 2004). This program implements an MCAT procedure based on the conceptual framework of Reckase (2009) and Thompson & Weiss (2011), utilizing the MIRT-3PL model.

Test takers can access the test via the website. The website interface includes the Homepage, Test Registration, Staff, and User Manual. The MCAT program was developed by integrating digital competency domain content with the MIRT framework. The detailed procedures are as follows:

- *Item bank:* This study encompasses 211 items, categorized into three components: (1) media information and digital literacy (70 items), (2) skill for using, developing, and problem-solving with digital tools (102 items), and (3) adaptive digital transformation (39 items). The parameters for the test items include (*a*) a discrimination value ranging from 0.49 to 5.19, (*b*) a difficulty value between -3 and 3, and (*c*) a guessing value from 0 to 0.3. The EAP reliability score was found to be 0.865, which is considered a suitable value (Baker & Kim, 2017; Lunz et al., 1990)

- *Initial item selection:* The MCAT program establishes the initial testing criteria by selecting a first item with a medium level of difficulty parameter (*b*). This study did not use prior information about the examinee's ability; therefore, the program selects the initial question with *b* close to 0. This approach offers a balanced starting point and enables the adaptive process to proceed efficiently (Hambleton & Swaminathan, 1985; Thompson & Weiss, 2011).

- *Ability estimation (θ):* In MCAT, the ability estimation used is based on MIRT. Appropriate estimation methods include Maximum Likelihood Estimation (MLE) (Lord & Novick, 2008), Expected A Posteriori (EAP), and Maximum A Posteriori (MAP) methods. Particularly, when answer patterns consist entirely of correct or incorrect answers, EAP and MAP methods are preferred due to their robustness and lower standard errors (Hambleton & Swaminathan, 1985). Ability estimation (θ) in this MCAT system was conducted using Bayesian statistics, specifically the MAP estimation method to estimate examinees' abilities.

- *Item selection*: The process is designed to optimize the accuracy of ability estimation by reducing the variance of the multidimensional ability vector (Frey & Seitz, 2009). In this study, the Maximum Likelihood (ML) method was employed in combination with the Maximum Information (*mi*) Criterion, matching each item's information to the examinee's current ability level (θ). This MCAT system approach selects items that provide maximum information at the point where $\theta = m_i$, with m_i derived using the MIRT-3PL model (Embretson & Reise, 2013). However, selecting items only based on the closeness of the b_i to θ may sometimes result in items that are too difficult for examinees. Therefore, consideration of item characteristics under the 3PL model is essential. Both ML and Bayesian Fisher Information (BF) methods are commonly applied for item selection and are compatible with various ability estimation strategies (Reckase, 2009). In the initial stage, a single item is chosen from the moderate-level module to begin. In subsequent stages, items are selected item-by-item estimation of the examinee's ability (θ). Selecting items only in the initial stage, a single item (N=1) is chosen from the moderate-level module to begin. In subsequent stages, items are selected based on item-by-item estimation of the examinee's ability (θ) based on cyclic content-balancing, the following formula: $k = ((N-2) \bmod 3) + 1$ where $k \in \{1,2,3\}$. Step 2 (N=2): Selects an item from Dimension 1. Step 3 (N=3): Selects an item from Dimension 2. Step 4 (N=4): Selects an item from Dimension 3. Step 5 ($N \geq 5$ onwards): Cycles back to Dimension 1 and continues this 3-step rotation sequentially until test termination.

- *Termination criterion:* The final step is termination criterion. In general, there are two primary rules for terminating a test: (1) the Fixed-Length Rule, where the test ends after a predetermined number of items, and the Variable-Length Rule, where the number of items varies (Reckase, 2009); and (2) the Standard Error of Estimation: SE(θ), exceeds a threshold for the standard error of ability estimation. For MCAT research, it is commonly found that the termination criterion is based on an SE(θ) set at 0.3 (Eggen & Straetmans, 2000). This MCAT system's performance is examined using RMSE and average bias (Hodson, 2022).

The study employed the Conditional Sympson-Hetter method for item exposure control. Examinees were grouped by ability levels. This mechanism was implemented to prevent specific items from being overexposed exclusively to high- or low-ability examinee groups.

Procedures

Phase 1 involved the development and quality validation of the item bank. We developed an item bank for a digital competency assessment based on three Components: (1) information

and digital media literacy, (2) skills in using, developing, and problem-solving with digital tools, and (3) adapting to digital transformation. The content validity was verified by five experts —three in information technology and two in educational measurement and evaluation. The item bank was administered to 1,256 undergraduate students; all participants completed the test and were included in the calibration analysis. The items were evaluated using a MIRT-3PL model. The cutoff score was determined using a criterion zone on the wright map to categorize examinees as either pass or fail.

Phase 2 involved the development and effectiveness validation of the MCAT program. We developed an MCAT program based on the concepts proposed by Reckase (2009) and Thompson and Weiss (2011). This program was divided into three steps: (1) studying the components of the MCAT system, (2) designing and developing the testing system, and (3) validating the quality of the MCAT system. The system was piloted with 112 undergraduate students; all completed the system trial and were included in the analysis. Each participant completed the test within 50 minutes via the online platform. The testing system evaluation was conducted in two parts.

First, the content validity of the system evaluation form was verified using the IOC, based on evaluations by seven experts in information technology. All experts held at least a master's degree and had at least five years of relevant professional experience. Second, test-takers' satisfaction regarding system usability was assessed immediately after completing the test.

Data Analysis

The quality of the digital competency test was assessed through construct validity using multidimensional analysis. Item parameter estimation and model fitting for the MIRT-3PL model were performed using the mirt package in the R software. Item parameters were estimated using the Marginal Maximum Likelihood estimation via Expectation-Maximization (MML-EM) algorithm. No missing data imputation was required. Item-fit statistics, including Infit and Outfit mean square (MNSQ) values, were examined. The reliability of the test was evaluated using multidimensional analysis with the Expected A Posteriori (EAP) reliability coefficient. The structural validity of the model was verified through confirmatory factor analysis. Furthermore, the efficiency of the MCAT program was examined using the RMSE and average bias to evaluate the accuracy of examinees' ability estimation.

Findings

Development of the Item Bank

Analysis of the 240 test items using the MIRT-3PL model showed that 211 items (87.92%) met the established criteria. Using Baker and Kim's (2017) criteria—discrimination parameters greater than 0, difficulty parameters ranging from −3 to +3, and guessing parameters not exceeding 0.3—the items that passed demonstrated an average discrimination (*a*) of 1.44, an average difficulty value (*b*) of 0.24, and an average guessing value (*c*) of 0.13. Although the maximum discrimination reached 5.19, reflecting exceptional diagnostic precision near the cut-off threshold, further inspection revealed a low standard error (SE = 0.039) and no model misfit, confirming parameter stability. These findings are consistent with the recommendations of Baker and Kim (2017), who suggested that high-quality test items should have difficulty parameters between −3 and +3, as the item characteristic curve (ICC) can effectively capture variations

across examinee ability levels. In addition, the item bank was categorized into three modules based on the difficulty parameter (*b*). Most items were of medium difficulty (−1<*b*<1), totaling 115 items, followed by difficult items (1<*b*<3) with 60 items, and easy items (−3<*b*<−1) with 36 items. Weiss (2013) indicated that an effective adaptive testing program should include at least 100–200 items in its item bank. Although the present item bank meets this recommendation, its development represents an initial stage, and future research may expand and refine the item bank to support more comprehensive adaptive testing applications, as presented in Table 4.

Table 4

Item Quality and Test Parameters

Components	Discrimination	Difficluty (-3≤*b*≤3)			Guessing	Item
		Easy	Medium	Difficult		
1. Media information and digital literacy	0.50 to 2.44	-3.00 to -1.04	-0.99 to 0.85	1.03 to 3.00	0 to 0.30	70
2. Skill for using, developing, and problem-solving with digital tools	0.50 to 5.19	-2.72 to -1.03	-0.89 to 0.89	1.04 to 2.97	0 to 0.30	102
3. Adaptive digital transformation	0.49 to 2.48	-2.04 to -1.07	-0.38 to 0.94	1.06 to 1.92	0 to 0.30	39
	a_{ave} = 1.44		b_{ave} = 0.24		c_{ave} = 0.13	211

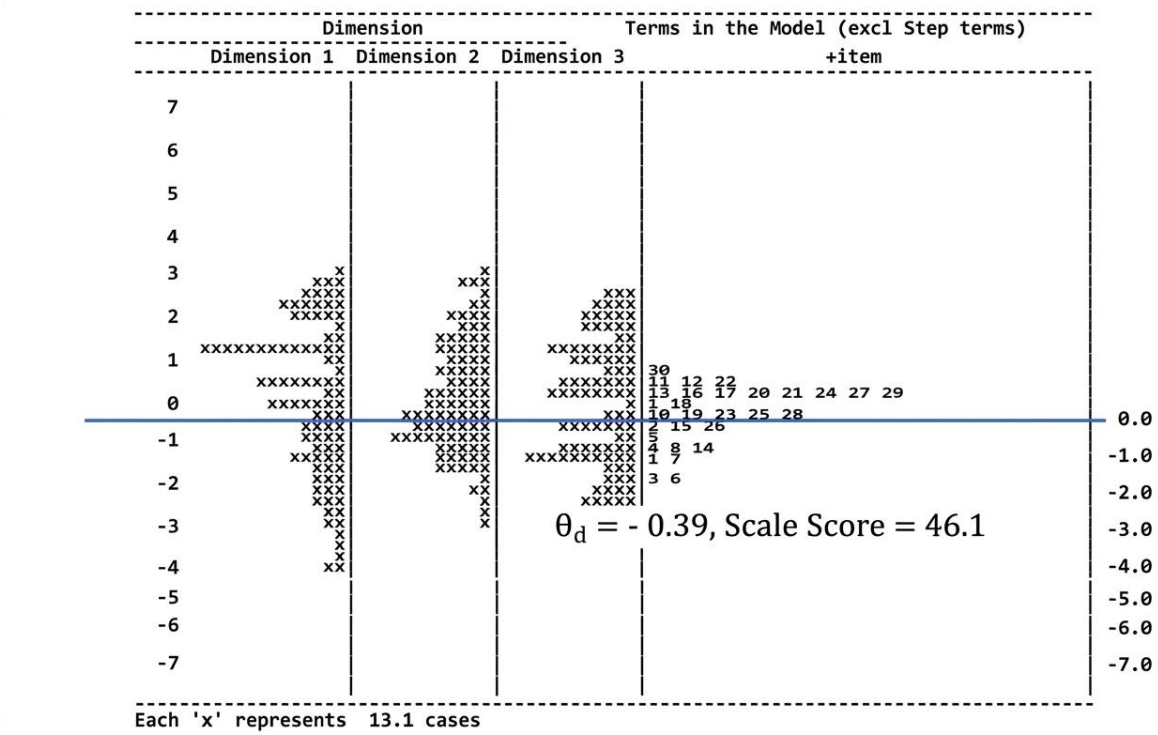
The examinees' responses indicated that the test aligned best with the three-parameter logistic model, as it presented lower G^2, AIC, and BIC indices than both the Rasch model and the 2 PL model (G^2 = 3429.35, 3542.04, and 3470.00; AIC = 4931.13, 4975.83, and 4961.79; BIC = 5045.23, 5066.79, and 5137.86, respectively) (Briggs & Wilson, 2003). A model fit comparison among the unidimensional, consecutive unidimensional, and between-item multidimensional model revealed that the between-item multidimensional model provided the best fit to the data, as evidenced by lower deviance index (G^2), AIC, and BIC values (G^2 = 44078.87, 42581.74, and 40834.06; AIC = 44140.87, 42649.74, and 40906.06; BIC = 44174.94, 42687.11, and 40945.63, respectively) (Briggs & Wilson, 2003). Statistically significant differences were observed at the .05 level. ($\chi^2_{Uni,between\text{-}item}$) = 64.89, p< .00 and ($\chi^2_{con,between\text{-}item}$) = 102.77, p< .00, respectively). Item-fit indices were evaluated using two fit statistics: Outfit MNSQ and Infit MNSQ, with values between 0.60 and 1.50 indicating an acceptable fit to the model (Lunz et al., 1990). The results indicated that Outfit MNSQ values ranged from 0.62 to 1.05, and Infit MNSQ values ranged from 0.74 to 1.04, all within acceptable threshold criteria. All statistical analyses were performed using the MIRT package and R software.

The digital competency cutoff score was determined by defining a criterion zone on the Wright map using ACER ConQuest with maximum likelihood (ML) estimation. The test defined cut-off thresholds at two levels: scores above the threshold indicated meeting the criterion, whereas scores below the threshold indicated not meeting the criterion. scores above the threshold position indicated passing the criteria, and scores below the threshold position indicated failing the criteria. The cutoff score, defined as the mean ability threshold (θ_d), was calculated from the average item difficulty categorized by dimension. The corresponding scale score was computed using the formula 50 + 10(θ_d) (Junpeng et al., 2018). The results indicated

that the overall digital competency test has a cut-off score at the mean ability threshold (θ_d) of -0.39, the scale score is 46.1, illustrated in Figure 2.

Figure 2

Cutoff Scores on the Wright Map of Digital Competency Levels



The digital competency of undergraduate students was assessed under the National Qualifications Framework for Higher Education using the criterion-zone method plotted on a Wright map to determine the cutoff score. The overall digital competency test had a cutoff score at the mean ability threshold (θ_d) of -0.39. The corresponding scale score was 46. The results indicated that 69.27% of students met the competency criterion, with the highest levels of digital competency in Component 1, followed Component 3, and the lowest level in Component 2. The findings on digital competence are consistent with those of previous studies. Techataweewan and Prasertsin (2018) examined four skills of digital literacy, including operational skills, thinking skills, collaboration skills, and awareness skills. Li and Ranieri (2010) identified three dimensions of digital competency: technology, cognitive domain, and ethics or social relationships, as summarized in Table 5.

Table 5

Digital Competency at the Undergraduate Level

Digital Competency	θ_d	scale score	Passed *N*	Not passed *N*	Total *N*
1. Information and Digital Media Literacy	-1.03	40	989 (78.74 %)	267 (21.26%)	1,256 (100%)
2. Skill for using, developing, and problem-solving with digital tools	-0.04	50	610 (48.57%)	646 (51.43%)	1,256 (100%)
3. Adaptability Digital Transformation	-0.09	50	737 (58.68%)	519 (41.32%)	1,256 (100%)
Overall Digital Competency	-0.39	46	870 (69.27%)	386 (30.73%)	1,256 (100%)

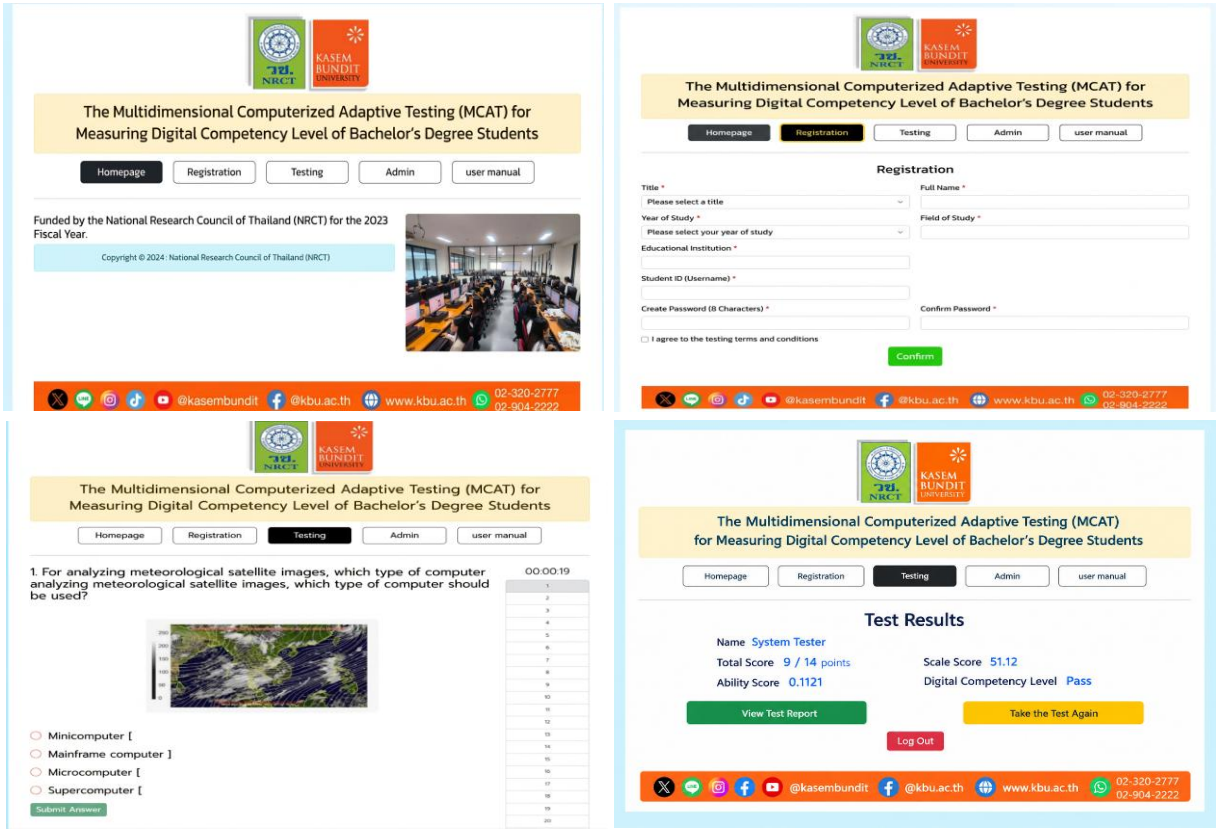
Note. θ_d the mean ability threshold, scale score 50 + 10(θ_d), *N* number of examinees

Development of the MCAT Program

The development of the MCAT program resulted in an online testing system built using PHP and a MySQL database. The system was implemented as a web application (https:// www.digitalmcat.com/) for assessing digital competency at the undergraduate level. The program interface consists of a homepage, test registration page, testing page, admin, and user manual, as shown in Figure 3. The program is divided into two sections: the tester section and the administrator section. In the tester section, users must register by providing their title, first and last name, academic year, field of study, institution, password, and password confirmation. After registration, the system directs them to the test page. Test takers must complete the test in the order presented on the screen and cannot return to change previous answers. Test takers who cancel before completing the test will not be able to have their scores processed. The MCAT system for measuring multiple digital competencies consists of the five steps previously described. Item bank: the bank comprised 211 items. The cutoff score was determined using the criterion zone method on the Wright map, with a single threshold separating the digital competency levels of 'passed' and 'not passed'. The cutoff score was established based on the mean ability estimate (θ_d) of -0.39. Starting the first item: the test begins with an item of medium difficulty (b = 0). When the initial value of the examinees' latent trait estimate was unknown, it was assumed to be 0 (Thompson & Weiss, 2011). Ability Estimation (θ): The Bayesian updating estimation method was employed to estimate examinees' ability. Item selection: items were selected using the maximum information method, prioritizing items whose difficulty levels were closest to the estimated θ from the previous item (Embretson & Reise, 2013). Termination Criterion: the criterion was set as the variable-length test with SE(θ) < 0.3 and mean ability estimates (θ) of 0.08. The number of administered items ranged between 4 and 24 with an average of 20 items. Kuo et al. (2015) found that the MCAT program is an efficient and acceptable method for assessing knowledge and skills across multiple dimensions.

Figure 3

Format of the MCAT Program



The ability of the MCAT system to predict student performance using the test items was determined through the item bank. The MCAT system has a personal data security system with registration, password creation, and identity verification. The comparative performance of the MCAT system under two criteria—(a) a fixed-length test of 30 items and (b) a variable-length test with SE(θ) < 0.3—is presented in Table 6. The results are summarized as follows: (1) the estimated ability (θ) ranged from -2.74 to 2.83, with comparable mean values under both criteria (0.12 and 0.08); (2) scale scores ranged from 22.56 to 78.29, showing similar mean scores (51.08, 50.79); (3) the mean test score varied considerably (9.86, 19.53) because the maximum total scores differed based on the termination criteria; and (4) the mean number of items differed between the criteria, with criterion (a) requiring more items than criterion (b) (30 items, 20 items). However, the number of subjects who passed the exam was similar under both criteria. The key advantage of criterion (b) is its fewer items, resulting in a shorter testing time compared to criterion (a).

Table 6

Student Trial Results with the MCAT System

Ability	termination criteria	*Min*	*Max*	*M*	*SD*
Theta (θ)	(a) 30 items	-2.74	2.83	0.12	0.78
	(b) SE(θ) < 0.3	-2.68	2.83	0.08	0.77
Scale Score	(a) 30 items	22.56	78.29	51.08	7.88
	(b) SE(θ) < 0.3	23.20	77.99	50.79	7.66
Test Score	(a) 30 items	1	18	9.86	3.41
	(b) SE(θ) < 0.3	4	28	19.53	3.90
Number of Item	(a) 30 items	30	30	30	0
	(b) SE(θ) < 0.3	4	30	20	4
N	(a) 30 items	Passed = 90		Not Passed = 22	
	(b) SE(θ) < 0.3	Passed = 88		Not Passed = 24	

Note. *Min* minimum, *Max* maximum, *M* mean, *SD* standard deviation, *N* number of examinees

Validity of the MCAT Program

The results of the performance comparison in ability estimation, under different test termination criteria revealed that: (a) a variable-length test with SE(θ) < 0.3 showed significantly different RMSE and bias values compared to (b) a fixed-length 30-item test at .01 significance level (F=32.484, sig < .001). These performance data are summarized in Table 7.

Table 7

Comparison of Performance of Ability Estimation Methods

Dependent	Termination criteria	G^2	AIC	BIC	*M*	*SD*	Hotelling's Trace				
							Value	F	Hypothesis df	Error df	Sig
RMSE	(a) SE(θ) < 0.3	4.98	18.79	32.93	0.02	0.63	0.29	32.48	2.00	221	.000
	(b) 30 items	5.31	23.82	37.96	0.11	0.08					
Bias	(a) SE(θ) < 0.3				0.00	0.14					
	(b) 30 items				0.03	0.15					
Box's M =	246.277, F= 1.1154, sig = .077										

The performance of the MCAT program, summarized in Table 8, was evaluated using RMSE and bias values calculated from individual-level data ($\hat{\theta}j_{,}$). In this analysis, the true ability parameter (θ) could not be directly observed; therefore, the scores obtained from the full-length paper-and-pencil test were utilized as a criterion estimate (true ability). The analysis revealed that the accuracy of examinee ability estimation based on the termination criterion demonstrated that using a stopping rule of a variable-length test with SE(θ) < 0.3 was significantly better than that of (b) the fixed-length 30-item test (p < .01), within a 211 items bank. in Table 7. This research suggests that estimating the ability of examinees by the termination criterion of the variable-length test with SE(θ) < 0.3 is appropriate, indicating that the number of items ranged from 9 to 30 items, and the 22–item condition (RMSE = 0.112, bias = 0.079) revealed the highest overall accuracy in estimating examinees' abilities.

However, the authors acknowledge a methodological limitation in using full-length paper-and-pencil test scores as a baseline, as observed scores are inherently subject to measurement error and are constrained by test length, and therefore serve strictly as a criterion estimate rather than a pure reflection of the examinee's latent trait. Chen et al. (1998) suggest that the termination criterion can be defined by setting the test length to 10, 20, 30, 40, 50, or 60 items. The accuracy of the present study was supported by the low RMSE and bias values observed. This means that the estimated abilities were more accurate depending on the chosen termination criterion; the less error occurred, the smaller the differences in examinees' abilities. According to Sahin and Weiss (2015) the ability estimates in CAT for a 200-item bank can be used for certain purposes; however, the examinees' abilities should be considered along with other factors. Defining the termination criterion depended on the research conditions. According to Babcock and Weiss (2012), using the standard error of estimation as a termination criterion is more common than other criteria.

Table 8

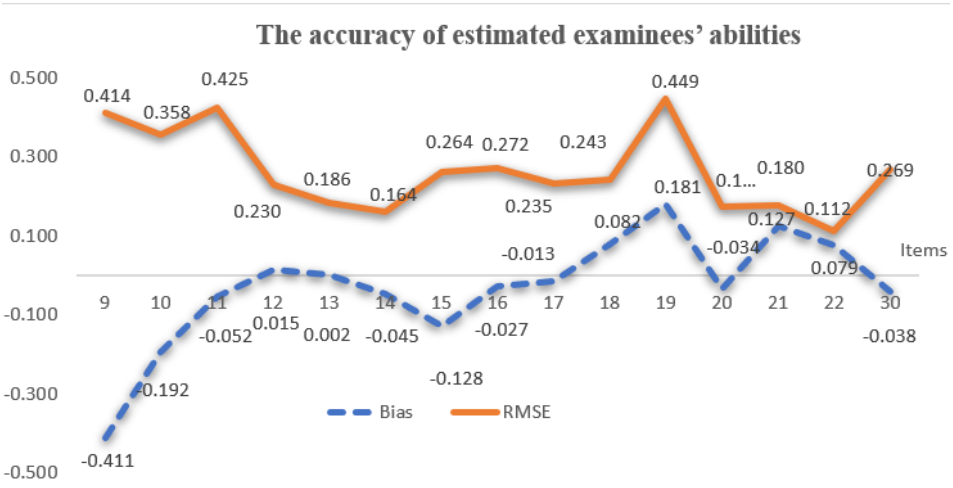
Accuracy Under Termination Criteria Based on SE(θ) < 0.3

Termination criteria	$\theta_{j,\mathrm{Paper}}$		$\hat{\theta}j_{,MCAT}$		Bias	RMSE
SE(θ) < 0.3	*M*	*SD*	*M*	*SD*		
9 items	-0.519	0.535	-0.930	0.596	-0.411	0.414
10 items	-0.849	0.502	-1.041	0.207	-0.192	0.358
11 items	-0.392	0.817	-0.444	0.419	-0.052	0.425
12 items	-0.355	0.794	-0.340	0.754	0.015	0.230
13 items	-0.098	0.494	-0.096	0.434	0.002	0.186
14 items	-0.111	0.564	-0.156	0.669	-0.045	0.164
15 items	0.112	0.456	-0.015	0.472	-0.128	0.264
16 items	0.490	0.659	0.463	0.611	-0.027	0.272
17 items	0.646	0.674	0.633	0.627	-0.013	0.235
18 items	0.317	0.428	0.399	0.367	0.082	0.243
19 items	0.357	1.452	0.537	1.063	0.181	0.449
20 items	0.657	0.304	0.623	0.190	-0.034	0.177
21 items	1.266	0.317	1.393	0.137	0.127	0.180
22 items	1.915	1.293	1.994	1.181	0.079	0.112
30 items	0.117	0.778	0.079	0.766	-0.038	0.269

Note. $\theta_{j,\mathrm{Paper}}$ is the criterion estimate (observed full-length paper test score) for individual *j*, serving as $\hat{\theta}_{j,\mathrm{MCAT}}$ is the estimated ability level for individual *j* calculated by the MCAT program

Figure 4

Estimated Accuracy of Examinees' Abilities



The evaluation of user satisfaction with the MCAT program revealed that test takers reported a high level of overall satisfaction (M = 3.95, SD = 0.93).

When examining individual items, the highest mean score was observed for "the program displays scores promptly" (M = 4.24, SD = 0.81). This was followed by "the program provides accurate scoring" (M = 3.97, SD = 0.79) and "the program meets users' needs" (M = 3.92, SD = 0.81), both of which reflected high levels of satisfaction. In contrast, the item with the lowest mean score was "the program allows users to easily correct errors," which demonstrated a moderate level of satisfaction (M = 3.10, SD = 0.72), as shown Table 9.

Table 9

Satisfaction Level of MCAT Program Users

Evaluation Items	*M*	*SD*	Level of Satisfaction
1. System Screen Features	3.45	0.72	High
1.1 Clear and readable font color and size	3.18	0.82	Medium
1.2 User-friendly and easy-to-understand toolbar	3.87	0.78	High
1.3 Clear and intuitive command icons	3.21	0.63	Medium
1.4 Appropriate placement of text and icons	3.54	0.83	High
2. Language and Information Aspects of the Program	3.68	0.96	High
2.1 The language used in the program is easy to understand	3.80	0.49	High
2.2 The terminology used is clear and comprehensible	3.77	0.92	High
2.3 The wording used is relevant and meaningful	3.27	0.89	Medium
2.4 The text and images presented in the test are clear	3.68	0.82	High
2.5 The number of answered is displayed during the test	3.90	0.89	High
3. Learning Aspects	3.75	0.91	High
3.1 The program is easy to use and not complicated	3.71	0.81	High
3.2 Clear instructions for submitting the examination	3.42	0.85	High

Table 9
(Cont.)

Evaluation Items	*M*	*SD*	Level of Satisfaction
3.3 The commands are easy to remember	3.89	0.82	High
3.4 The program can operate continuously	3.83	0.88	High
3.5 Help instructions assist users in understanding	3.91	0.62	High
4. Program Capabilities	3.78	0.53	High
4.1 The program displays scores promptly	4.24	0.81	Very High
4.2 The program provides accurate scoring	3.97	0.79	High
4.3 The program operates smoothly without interruption	3.67	0.57	High
4.4 The program allows users to easily correct errors	3.10	0.72	Medium
4.5 The program meets users' needs	3.92	0.81	High
5. Overall Satisfaction	3.95	0.93	High

Discussions

Development of the Item Bank

To select high-quality items for inclusion in the item bank, the study utilized the MIRT-3PL model. The item bank retained items featuring discrimination parameters (*a*) greater than 0, difficulty parameters (*b*) ranging from -3 to +3, and guessing parameters (*c*) not exceeding 0.30. These selection criteria aligned with Baker and Kim's (2017) recommendation that quality items should maintain difficulty values within $-3<b<3$, allowing the item characteristic curve (ICC) to effectively capture the variance across the entire spectrum of examinee ability parameters.

Development of the MCAT Program

The MCAT program demonstrates high measurement precision in estimating examinees' ability levels (θ) by employing a maximum information item selection mechanism, with discrepancies from the true ability not exceeding 0.062. This finding aligns with Embretson and Reise (2013), who noted that under a three-parameter model, selecting items based on maximum information at an examinee's current ability level ensures the selection of items that maximize test information while accounting for the impact of the guessing parameter.

For the test termination criterion of a variable-length test program, $SE(\theta) < 0.3$, a test bank of 22 items yielded the best accuracy in estimating test-takers' abilities, as it had the lowest RMSE for a test bank of 211 items. This result aligns with Homjan et al. (2016), who demonstrated that an item bank of 200 items provides optimal accuracy in examinee ability estimation. Furthermore, it supports the recommendation by Babcock and Weiss (2012) that stopping rules may incorporate multiple standard error thresholds, provided they are evaluated in conjunction with the accuracy and reliability of the estimated ability parameters.

The evaluation of user satisfaction with the MCAT program revealed that test-takers reported a high level of overall satisfaction. Regarding individual items, the three with the highest mean scores were "prompt display of scores," "accurate scoring," and "fulfillment of users' needs." These findings are consistent with prior research evaluation of user satisfaction with information systems in terms of accuracy and relevance to user needs, operational efficiency, ease of use, system design convenience, and data security.

Conclusions

Following a thorough evaluation, experts confirmed the validity of the MCAT program and highlighted its strong performance across all assessed dimensions, including content accuracy, indicator quality, and program structure. The program structure comprises several key components: development of an item bank based on the MIRT-3PL model; determination of the cutoff score using the Wright map; initial item selection at a medium difficulty level ($b = 0$); ability estimation (θ) using the Bayesian updating method; subsequent item selection using the maximum-information (mi) method; and a variable-length termination criterion with $SE(\theta) < 0.3$. Each component demonstrated strong performance and contributed to the overall effectiveness of the MCAT program.

In addition, the MCAT program investigated the accuracy of ability estimation through a pilot study involving undergraduate students. The results demonstrated outstanding usability, accessibility, and flexibility, indicating its high practicality for diverse educational contexts. The testing process is designed to be easy for students to navigate, and compatibility with various devices and browsers ensures accessibility and convenience for users with diverse abilities. Moreover, the adaptive testing approach requires less time and fewer items than traditional tests, with the optimal number of items for accurate ability estimation ranging from 9 to 30, and 22 items. yielding the highest accuracy. Overall, the MCAT program offers a well-balanced and user-centered assessment characterized by flexibility and ease of use, making it suitable for equitable implementation across large, medium, and small universities.

Suggestions

The MCAT system effectively measures the digital competence of undergraduate students across all disciplines because questions are based on a comprehensive conceptual framework, namely the Higher Education Qualification Standards 2022, DigComp 2.0 by the European Commission, 21st Century Learning Skills, and the Top 10 Skills identified by the World Economic Forum. The MCAT system covers digital competence across three key Components: media information and digital literacy; Skill for using, developing, and problem-solving with digital tools; and adaptive digital transformation.

For future system development, and to enhance user confidence and system stability, future iterations of the MCAT platform should implement an auto-saving and crash-recovery feature. Since adaptive testing constraints restrict examinees from navigating back to previous items—which has resulted in lower user satisfaction regarding error correction—this automated logging feature will continuously preserve examinee progress and seamlessly restore responses in the event of technical disruptions or system failures.

Furthermore, for future research, we recommend integrating machine learning (ML) concepts, such as decision regression trees (DRT) and random forest (RF), in coordination with the MCAT framework, leveraging current artificial intelligence (AI) capabilities to achieve optimal predictive accuracy. These predictive models can be integrated into an Early Warning System (EWS) to evaluate student performance in digital competencies. By tracking learning analytics and quiz histories, educational institutions can proactively implement academic support and tailored remedial coursework before students fall behind.

Author Contributions

ASAT: Conceptualization, methodology, construct and validation instrument, data collection, analysis, writing of the original draft, software design, and funding acquisition.
KAMT: Conceptualization, validation, review, and editing, project administration, and supervision.

Declaration of the Use of Generatve AI

The authors utilized Gemini 3.6 Flash (Free tier) exclusively for language refinement, grammar correction, and readability enhancement during the drafting of this manuscript. All AI-suggested modifications were thoroughly evaluated and edited by the authors, who retain complete responsibility for the final published content.

Ethics

This study received ethical approval from The Ethical Committee of Kasem Bundit University, Thailand, Approval No R001/67P.

Funding

This work was supported by the National Research Council of Thailand (NRCT) under Grant number N42A660894.

References

Babcock, B., & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement? Journal of Computerized Adaptive Testing, 1(1), 1-18. https://doi.org/10.7333/1212-0101001

Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R* (Vol. 10). Springer. https://doi.org/10.1007/978-3-319-54205-8

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*(1), 87-100.

Brossman, B. G., & Guille, R. A. (2014). A comparison of multi-stage and linear test designs for medium-size licensure and certification examinations. *Journal of Computerized Adaptive Testing*, *2*(1-4), 18-36.

Chen, S. K., Hou, L., & Dodd, B. G. (1998). A comparison of maximum likelihood estimation and expected a posteriori estimation in CAT using the partial credit model. *Educational and Psychological Measurement*, *58*(4), 569-595. https://doi.org/10.1177/0013164498058004002

Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological measurement*, *60*(5), 713-734. https://doi.org/10.1177/00131640021970862

Elliott, G. (2004). *Global business information technology: An integrated systems approach*. Pearson Education.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press. https://doi.org/10.4324/9781410605269

Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis—based models. *Applied Psychological Measurement*, *34*(1), 10-26. https://doi.org/10.1177/0146621609336112

Fraumeni, B. M., & Liu, G. (2021). Summary of World Economic Forum,"The Global Human Capital Report 2017—Preparing people for the future of work". In *Measuring human capital* (pp. 125-138). Academic Press. https://doi.org/10.1016/B978-0-12-819057-9.00008-1

Frey, A., & Seitz, N. N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, *35*(2-3), 89-94. https://doi.org/10.1016/j.stueduc.2009.10.007

Hambleton, R. K., & Swaminathan, H. (1985). Estimation of item and ability parameters. In *Item response theory: Principles and applications* (pp. 125-150). Springer. https://doi.org/10.1007/978-94-017-1988-9_7

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions, 15*(4), 5481-5487. https://doi.org/10.5194/gmd-15-5481-2022

Homjan, S., Thayreakum, S., & Kumpon, S. (2016). The efficiency of a Computerized Adaptive Testing (CAT) by applying multidimensional item response models. *Social Sciences Research and Academic Journal*, *11*(31), 73–88.

Junpeng, P., Inprasitha, M., & Wilson, M. (2018). Modeling of the open-ended items for assessing multiple proficiencies in mathematical problem solving. *The Turkish Online Journal of Educational Technology*, *2*, 142-149.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Li, J., Sun, J., Shao, M., Lai, Y., & Chen, C. (2025). A new multidimensional computerized testing approach: On-the-fly assembled multistage adaptive testing based on multidimensional item response theory. *Mathematics*, *13*(4), 594. https://doi.org/10.3390/math13040594

Li, Y., & Ranieri, M. (2010). Are 'digital natives' really digitally competent?—A study on Chinese teenagers. *British Journal of Educational Technology*, *41*(6), 1029-1042. https://doi.org/10.1111/j.1467-8535.2009.01053.x

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, *3*(4), 331-345. https://doi.org/10.1207/s15324818ame0304_3

Piton-Gonçalves, J., & Aluísio, S. M. (2012, October). An architecture for multidimensional computer adaptive test with educational purposes. *In Proceedings of the 18th Brazilian symposium on Multimedia and the web (pp. 17-24).* Association for Computing Machinery. https://doi.org/10.1145/2382636.2382644

Reckase, M. D. (2009). Historical background for multidimensional item response theory (MIRT). In *Multidimensional item response theory* (pp. 57-77). Springer. https://doi.org/10.1007/978-0-387-89976-3

Sahin, A., & Weiss, D. J. (2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory and Practice*, *15*(6), 1585-1595.

Shao, M., Sun, J., Li, J., Wang, S., & Lai, Y. (2024). Multidimensional hybrid computerized adaptive testing based on multidimensional item response theory. *IEEE Access*, 169079-169101. https://doi.org/10.1109/ACCESS.2024.3492188

Techataweewan, W., & Prasertsin, U. (2018). Development of digital literacy indicators for Thai undergraduate students using mixed method research. *Kasetsart Journal of Social Sciences*, *39*(2), 215-221. https://doi.org/10.1016/j.kjss.2017.07.001

Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, *16*(1).

Vuorikari, R., Punie, Y., Gomez, S. C., & Van Den Brande, G. (2016). DigComp 2.0: The digital competence framework for citizens. Update phase 1: The conceptual reference model (No. JRC101254). Publication Office of the European Union.

Wang, C., & Chang, H. H. (2011). Item selection in multidimensional computerized adaptive testing—gaining information from different angles. *Psychometrika*, *76*(3), 363-384. https://doi.org/10.1007/s11336-011-9215-7

Wang, C., Weiss, D. J., Su, S., Suen, K. Y., & Basford, J. (2022). Multidimensional computerized adaptive testing: A potential path toward the efficient and precise assessment of applied cognition, daily activity, and mobility for hospitalized patients. *Archives of Physical Medicine and Rehabilitation*, *103*(5), S3-S14. https://doi.org/10.1016/j.apmr.2022.01.002

Weiss, D. J. (2013). Item banking, test development, and test delivery. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 185–200). American Psychological Association. https://doi.org/10.1037/14047-010