DEVELOPMENT OF ELECTRONIC ITEM BANK FOR MEASURING THE LEARNING OUTCOMES OF STUDENTS IN UNDERGRADUATE EDUCATION PROGRAMS

Jiraporn Meesanga, Kamonwan Tangdhanakanond*, and Shotiga Pasiphol

Faculty of Education, Chulalongkorn University

*Corresponding Author/Jiranid@outlook.co.th

ABSTRACT

The objectives of this research were 1) to develop an electronic item bank for measuring the learning outcomes based on the Thai Qualifications Framework for Higher Education (QTF: HEd) of undergraduate students in education programs and 2) to assess the efficiency of the electronic item bank. The research sample consisted of 1,631, obtained through multi-stage sampling. The statistics used in this research were the following: descriptive statistics, discrimination parameter, difficulty parameter, OUTFIT MNSQ, INFIT MNSQ, Deviance Statistic (G^2), Akaike information criterion (AIC), and EAP reliability. The results of the research were as follows: 1.) The electronic item bank consisted of five components; user management, item management, examination management, evaluation management, and scoring management. The electronic item bank has the mean of item difficulty parameter of 0.069 and the mean of item discrimination parameter of 0.862. The test randomized from the electronic item bank has the outfit MNSQ values ranging from $.75$ to 1.19, and the infit MNSQ values ranged from $.85$ to 1.15. The EAP reliability was 0.707. The structural validity test revealed that the multidimensional approach was a better fit than the composite approach and consecutive approach, respectively. The cut-off score of the learning outcome obtained by analyzing the criterion zone on the Wright map was dichotomous: pass and fail. 2.) The evaluation of the electronic item bank by experts revealed that the overall quality of the system was at the highest level. Additionally, the students were highly satisfied with the system.

Keywords: Thai Qualifications Framework, Multidimensional Item Response Theory, validity and reliability of measurement.

Introduction

The change of global society has affected Thai society. Moreover, Thai society is facing rapidly changing economic, social, political, technological, energy, and environmental situations. As a result, Thailand has built various policies to drive the country forward on par with other countries. In particular, higher education management is considered as a response to solving critical problems and guiding the sustainable development of the nation and the locality by accelerating to strengthen the immunity in the country under the philosophy of sufficiency economy. Additionally, it must promote the development of the country to be able to compete in the ASEAN community and the Global community, which gives importance to the development of people and Thai society for quality. All higher education institutions have to upgrade the quality of their programs, enhance the quality of graduates to be accredited at ASEAN and global levels. Therefore, the Ministry of Education has the policy to reform education by establishing the Thai Qualifications Framework for Higher Education (TQF: HEd) as a framework for the country's qualification system in higher education to help systematically set standards in educational management at every stage by providing opportunities for higher education institutions to organize courses as well as a variety of teaching and learning processes. This is to ensure the quality of graduates who will meet the standard of learning outcomes that are expected to be able to pursue careers with happiness and pride. The satisfaction of graduate users and being a good person in society helps to strengthen and develop the capability of Thailand. (Office of the Higher Education Commission, 2009)

The bachelor's degree in education programs is of greatest importance in raising the quality of teachers which will contribute to education in the development of the country. Because teachers are responsible for providing education for the country's citizens, the country has good citizens as an important force in driving economic and social development affecting the further progress of the country. Currently, it is found that the teacher production system has problems. For instance, the teacher production institutes have very different standards, the curriculum of each institute does not emphasize the integration of new teaching competencies and the respective government agencies that are responsible for examining the quality of curriculum standards of teacher-producing institutions lack rigor and seriousness, etc. (Rohitasthira, 2016). Therefore, the quality of graduate teachers should be measured and evaluated to ensure that they are consistent with the graduate learning outcomes of the Thai Qualifications Framework for Higher Education (TQF: HEd) to ensure the quality of teachers. At present, there are no specific tools to measure and evaluate learning outcomes.

Existing measuring tools are knowledge, which is a form of knowledge test in each subject that can measure learning outcomes in this area very well. The moral and ethical aspect has several tools to measure this for teachers that can be applied. In addition, the measurement of morals and ethics should be depended on observing relevant behavioral expressions to measure more accurately.

Therefore, the researcher does not aim to develop a tool to measure learning outcomes in such areas. However, it aims to develop tools to measure learning outcomes for various skills, including cognitive skills, interpersonal skills and responsibilities, numerical analysis skills, and communication and information technology skills. As previously mentioned, all skills are essential for 21st-century learners that enable students to successfully live in today's world. There should be specific and quality measuring instruments for assessing such skills. Learning management skills is considered a basic competency for the practice of teaching professions. Despite some instruments for evaluation and assessment when students practice teaching in schools to increase the reliability of the measurement and evaluation process, there should be a measurement instrument based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs and additional standards for assessing these skills. For this reason, the researcher is interested in developing the quality of measurement instruments that provide accurate and useful information. This measuring instrument has psychometric properties in terms of validity, reliability, and good item statistics (Friedenberg, 1995), in which the model for measuring learning outcomes relates to a psychological variable. Variables are typically made up of sub-dimensions or sub-elements and tests that measure latent trait or ability of the test taker to these variables often have a structure of properties within complex variables. This means that test-takers will need more than one ability to answer these tests. Therefore, the multidimensional latent traits are valid and appropriate properties for both analyzing and interpreting the latent trait or ability of the test taker. To conduct a proper analysis, Multidimensional Item Response Theory Model: MIRT Model should be used (Kanjanawasee, 2012).

Nowadays, computers used in examinations are convenient, economical, and quick in conducting tests and collecting data, as well as making the effective management of the examination. The researcher aimed to develop the electronic item bank for measuring the learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs using computerized testing. The exams of the measurements are stored in the electronic item bank with the management system that can be a source of exam collection and can use the exam systematically. The electronic item bank consisted of items for 4 domains; 1) cognitive skill, 2) interpersonal skill and responsibility, 3) numerical analysis, communication, and information technology skills, and 4) learning management skills. Any associated party was able to access the item bank and acquired a set of items for their educational measurement and evaluation. Accurate measurement and evaluation were crucial for planning and development in education.

Literature Review

Thai Qualification Framework for Higher Education (TQF: HED)

Thai Qualification Framework for Higher Education; TQF: HED is a tool for implementing policies to develop quality and educational management standards as stipulated in the National Education Act. National education standards and higher education standards can be concretely implemented in higher education institutions. The TQF: HED has clear guidelines for curriculum development, a change in teaching methods of teachers, student learning as well as measurement and evaluation of learning outcomes to ensure graduates meet the *academic achievement standards.* It focuses on the standard of graduate learning outcomes to guarantee graduate quality and communicate with relevant agencies and stakeholders to understand and be confident in the graduate process. It also aims to provide qualifications or degrees of any higher education institution in Thailand that is recognized and compared with good higher education institutions both domestically and internationally. (Office of the Higher Education Commission, 2009)

The standards of learning outcomes are set in the bachelor's degrees in education programs which are following the TQF: HED in 6 domains: (Office of the Higher Education Commission, 2009) 1) ethics and moral, 2) knowledge, 3) cognitive skill, 4) interpersonal skill and responsibility, 5) numerical analysis, communication, and information technology skills, and 6) learning management skill.

At present, there is no specific measurement and evaluation tool for measuring the learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs. This research aims to develop a tool to measure learning outcomes of student teachers in 4 domains; 1) cognitive skill, 2) interpersonal skill and responsibility, 3) numerical analysis, communication, and information technology skill, and 4) learning management skill. Domains 1-3 are essential skills for learners in the 21st century and will enable students to succeed live today. As for learning management skills, it is a primary qualification necessary for practice in teaching professions. Although there are some tools for educational measurement and evaluation, students practice teaching in schools. But to increase the reliability of the measurement and evaluation process, there should be a measurement tool that is consistent with the Thai Qualifications Framework for Higher Education (TQF: HEd) for undergraduate students in education and additional standards for assessing these skills. For this reason, the researcher is interested in developing an electronic item bank to measure learning outcomes according to the Thai Qualifications Framework for Higher Education (TQF: HEd) for undergraduate students in education programs.

When higher education institutions can organize courses as well as a wide range of teaching and learning processes, it can provide students with the standard of the expected learning outcomes.

This will enable students to become qualified graduate teachers which will lead to the development of the quality of education and further development of the country.

Multidimensional Item Response Theory Model: MIRT Model

Multidimensional Item Response Theory (MIRT) is developed in response to today's increasingly complex psychological and educational processes, making assessments more reliable. This theory can explain the relationship between test-takers traits and characteristics of the item. The test-takers traits include more than one ability to estimate parameters of the item and the test-takers. (Reckase, 2009) Therefore, the model used to estimate test takers' latent traits should be a multidimensional item response theory (MIRT Model). There are two types of MIRT models based on the concept of creating a different item.

Adams et al. (1997) define two different types of multidimensionality: between-item multidimensionality and within-item multidimensionality. The between-item multidimensionality occurs when the items are unidimensional, i.e., each item measures a single latent trait and each item is assigned to one and only one dimension. Several combinations of items, which are referred to as subtests or subscales, are designed to measure multiple dimensions, but each item belongs to only one dimension. The within-item multidimensionality occurs when items or a handful of them are multidimensional, i.e., each item is designed to measure simultaneously more than one dimension.
By creating an item structure to measure learning outcomes in this research, the measurement of the traits is relevant to the between-items multidimensionality, that each item only evaluated one trait at a time, but multiple latent traits could influence an individual success. Each influential latent trait was relevant to the other. According to MIRT Model, there is more than 1 latent trait influencing individual success on an item so it can be implied that there are 2 or more examinees' parameters. Therefore, considering the multidimensional ability of the test takers, it should help make them more consistent with the data. When the test taker uses systematically different multidimensional abilities to answer the test item, it correlates with the item difficulty parameter and item discrimination parameter and affects probability in answering the exam correctly. A multidimensional approach should be a better fit for data analysis (Kanjanawasee, 2012). Based on the literature review, MIRT Model should be fit for this research.

The Electronic Item Bank

The electronic item bank in this research is an item banking that has developed computer programs to have the ability to be a source of items collection, to support the management of items collection, and to use the systematic test for the assessment of learning outcomes per the Thai Qualification Framework for Higher Education (TQF: HED) of students in undergraduate education programs. At present, there is no electronic item bank of this kind. The electronic item bank has components that

are consistent with the components of the information system (Angsuchoti, 2019) as follows: 1) Hardware means a computer device used to collect data and process data to create information such as computers. In addition, information systems are stored in the network, which is a link between multiple computers together.; 2) Software is an instruction set or computer programs that are written to operate a computer.; 3) Stored Data is data that is collected in a computer system and will be run for processing by various applications.; 4) Personnel is an information system that will not be able to perform various tasks by itself if there is no one to manage them. The people here are (1) the user is generally the person who uses the information generated by the computer system; (2) The operator is generally the person who brings the information into the computer system, process the data and generate information, and wait for results from the computer system; and (3) the system operator and program developer. The system administrator is responsible for controlling the hardware system. For program developers, including personnel who are responsible for developing various applications to work for the computer, order processing and creating information in various work systems, and so forth.; 5) Procedures tell the user how to use the information from the computer system.

For the development of the electronic item bank in this research, it is the development of computer programs to be used as a data collection source, data processing of exams, quizzes, tests, and assessments for instructors to measure the learning outcomes of students in bachelor's degree in education programs. In addition, instructors can also select a test that matches the objectives of an examination.

The development of the electronic item bank has the same steps as the development of an information system. (Angsuchoti, 2019) as follows: 1) a survey for preliminary information on various issues related to the work system, i.e. the problem analysis, the *definition* of *scope* and requirement of the item bank, the development feasibility, the strategies that will improve the operational efficiency, and the cost estimation.; 2) Demand analysis is a survey of the user needs, usage in each area, the advantages and disadvantages of the current method of work to develop into a logical model.; 3) Detailed design of various parts of the system, including the display of results, data entry, retention processes, operations, and personnel related to the new system by starting the design from the device and computer program technology to be developed.; 4) Information systems development with programs and instruction sets so that the item bank can store the exams into the repository, select the exams from the repository for use, publish the exams, provide information about the exam repository and maintenance of security system of the exam repository.; 5) Procurement of equipment and instruction sets of the system to bring the equipment and components of the system to install and develop the new system of future demands.; 6) Testing the system before putting it into practice.; 7) Installation of the system and having to test the use of the new system to see if it is a function according to the intended purpose and design.; 8) Inspection by specialists bring the developed information system to the experts to evaluate the system.; 9) System

trial with the sample group.; 10) User satisfaction.; 11) Maintenance After the system has been installed and used.

The conceptual framework of the research

Thai Qualification Framework for Higher Education (TQF:	**Electronic item bank**	The efficiency of the electronic item bank:

Electronic item bank

Thai Qualification Framework for Higher Education (TQF:

The quality of items
- discrimination parameters(a)
- difficulty parameters (b)

Multidimensional Item Response Theory (MIRT)

The quality of tests
-The index of appropriate item
- EAP reliability
- Construct Validity by using MIRT

Setting the cut-off

The efficiency of the electronic item bank:
Before using
- utility
- feasibility
- propriety
- accuracy

After using
- service workability
- user experience
- user interface: visualization
- user interface: design
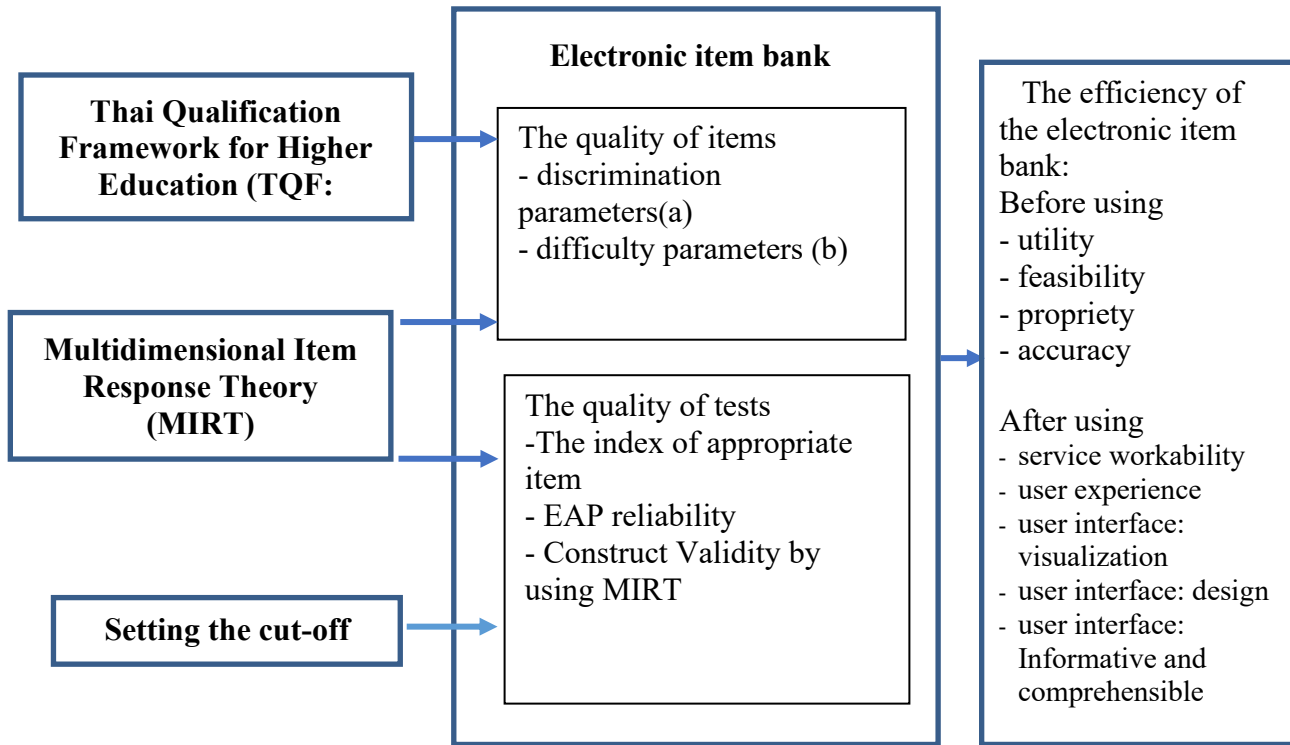- user interface: Informative and comprehensible

Figure 1: Research Framework

From the study of learning outcomes according to the National Higher Education Qualifications Framework of students in educational bachelor's degree programs. This makes the researcher interested in developing items and tests that are unique in the assessment of student learning outcomes based on the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs. The quality of items and tests is checked to obtain accuracy, appropriateness, validity, and reliability by applying the Multidimensional Item Response Theory (MIRT) and a cut-off score, which is defined for use in interpreting the tests. Thereafter, the quality of the items and tests is collected in an electronic item bank for learning outcomes under the TQF: HED for undergraduate students in education programs which higher education institutions can use to measure and assess the effectiveness of whether the graduate production has reached the expected quality or not. It can be used as evidence of educational quality assurance and the results obtained from the measurement and assessment can be used to create guidelines for further development of undergraduate students.

Methodology

The researchers employed a quantitative method to develop and evaluate the effectiveness of an electronic item bank for measuring learning outcomes based on the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs. The details are as follows.

Research design

It was considered the components of the developed electronic item bank, the quality of the items collected in the electronic item bank, and the quality of randomized tests from the electronic item bank. In addition, the cut-off scores for the test were set for interpreting learning outcomes under the Thai Qualification Framework for Higher Education (TQF: HED) of students in undergraduate education programs; and to assess the effectiveness of the electronic item bank.

Population

The participants in this research were 20,616 fourth-year undergraduate students for the second semester of the academic year 2019 in 51 autonomous universities or Rajabhat universities. (Source: Office of the Higher Education Commission, 2020)

Sampling is divided into two groups:

The first group was used to check the quality of the items before being stored in the electronic item bank as 4th-year undergraduate students in education programs. The sample size was determined using Yamane's Tables (Yamane, 1976, as cited in Vollakitkasemskul, 2011). Considering the sample size at the 99% confidence level when the discrepancy was 5%, the sample size was 869 people. Therefore, to strengthen the research and consider based on the feasibility of data collection, the researcher applied the scaling of the sample to 1,014 people, which were obtained by multi-stage random sampling.

The second group was used to examine the quality of a randomized test from the electronic item bank and check the quality of the electronic item bank for measuring learning outcomes based on the Thai Qualification Framework for Higher Education (TQF: HED) of 617 fourth-year undergraduate students in education programs. The sample size was determined by applying the sample size criteria of Comrey and Lee. Comrey and Lee (2016) stated that factor analysis should have a sample size of 500, which was very good, and should not be less than 200. Because the sample size was too small, it can lead to high errors and discrepancies in measurement, and low reliability. The sampling was chosen using the multi-stage random method.

Instruments in the research

1. The multiple-choice situational test was used in this research according to the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs, with a score ranging between 0 and 1.

2. The questionnaire was used to check the system of the electronic item bank by the experts and was the rating scale (very strongly agree, strongly agree, agree, strongly disagree, very strongly disagree) with 17 items for evaluating the efficiency of the electronic item bank in 4 areas: utility, feasibility, propriety, and accuracy.

3. The questionnaire was used to assess the undergraduate students' satisfaction with using the electronic item bank system and was the rating scale (very good, good, barely acceptable, poor, very poor) with 15 items, divided into 5 aspects, namely 1) service workability, 2) user experience, 3) user interface: visualization, 4) user interface: design and 5) user interface: informative and comprehensible.

Data Collection

1. The researcher and the experts created items to measure learning outcomes based on the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs that matched the operational definition, consisting of 4 areas: 1) cognitive skills, 2) interpersonal skill and responsibility, 3) numerical analysis, communication, and information technology skills, and 4) learning management skills. There was a total of 8 factors, namely 1) problem-solving ability, 2) intellectual leadership, 3) interpersonal relationships, 4) responsibility, 5) information technology ability, 6) communication ability, 7) learning management ability, and 8) measurement and evaluation of learner ability. It was a multiple-choice situational test, with a score of 0, 1, and a total of 342 items.

2. The created items were taken to check the content validity by experts. The researcher selected only the item with the IOC value between 0.60 – 1.00. Therefore, there were 330 passing items.

3. There was an examination of the quality of each item according to the multidimensional item response theory (MIRT). A total of 10 tests were collected and tested on the first group of 1,014 people, and the concurrent calibration method was used in the measurement.

4. Quality items were selected to be stored in the electronic item bank by considering the discrimination parameter according to the criteria of Baker and Kim (2017), eliminating items with negative discrimination parameters and selecting items with discrimination parameters of 0 or higher, and considering the difficulty parameters. The researcher accepted the items with the difficulty parameter between -3 and +3 as the quality items.

5. The researcher and the computer programmer jointly developed a computer program system, the electronic item bank for measuring learning outcomes according to the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in

education programs. The structure of the electronic item bank has been written, namely the database selection, program selection, security system, and computer programming instructions. There were also the main components in the operation of many computer programs in this structure of the electronic item bank.

6. The quality check of the system of the electronic item bank computer program was conducted by the 5 experts who had opinions on the system in 4 aspects, utility, feasibility, propriety, and accuracy.

7. The qualified items were stored in the electronic exam bank for measuring learning outcomes according to the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs.

8. The system of the electronic item bank has randomly arranged a test of learning outcomes according to the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs. The test contained a total of 78 items for quality checks.

9. The test was taken for a try-out with 617 students in the second group of samples 617 by using computer-based testing to test the quality as follows: 1) Construct validity by using the multidimensional analysis method; 2) The item fit statistics by using OUTFIT Mean Square and INFIT Mean Square; 3) The EAP reliability by using ConQuest program and undergraduate students to assess their satisfaction with using the electronic item bank system.

10. The student's test results were collected to determine the cut-off score by criterion zone on the Wright map from analysis by the ConQuest program.

11. The results of assessments of the students' satisfaction with using the electronic item bank system were collected and analyzed by using descriptive statistics.

Data Analysis

1. The quality of each item was analyzed according to the multidimensional item response theory (MIRT) considering the discriminant parameters and the difficulty parameters by analyzing the data in the R program before putting the item into the electronic item bank.

2. The analysis was done on the data to examine the quality of the test randomly acquired from the electronic item bank by using the ConQuest program.

3. The analysis was done on the data to determine the cut-off score of the learning outcome measurement according to the Thai Qualification Framework for Higher Education (TQF: HED) of undergraduate students in education programs by using the ConQuest program.

4. The analysis was done on the results of the quality assessment of the system of the electronic item bank computer program from experts and the results of the students' satisfaction

assessment for using the electronic item bank using descriptive statistics such as mean and standard deviation.

Research Ethics

The conduct of the research took into account the primary human research ethics, consisting of 3 principles: 1) Respect for the person, 2) Beneficence, and 3) Justice. We have requested permission to conduct a local study from the relevant agencies and ask for cooperation in providing information. We have received consent to participate in the research from the sample and maintained the confidentiality of the sample group.

Results of the research

1. The results of the development of the electronic item bank for measuring learning outcomes according to the Thai Qualifications Framework for Higher Education (TQF: HEd) for undergraduate students in education program consisted of: 1.1 Components of the developed electronic item bank 1.2 The quality of the items collected in the electronic item bank 1.3 The quality of the tests randomly drawn from the electronic item bank and the determination of the cut-off scores of the test as follows:

   1.1 Components of the developed electronic item bank:

       1.1.1 User management consists of logging in through the user's website and registering the user by dividing the user into two groups: the test taker and the administrator.

       1.1.2 Item management consists of a) Managing the details of the item, including information, questions, answers, indicators, components, skills to be measured which the administrator can add, edit, delete or turn off and on the status of the data; b) Conditions are determined in the preparation of the test set, such as taking the test from a random system, taking the test from the custom criteria, taking the test from the test set and ending the acceptance of the test; and c) Test set is a list of available items with test set codes, year of testing, created date, test set name, details of all items in that set, print out, delete, and end the test set as well as being able to add test sets from the system by randomly selecting the items to arrange the test sets according to the conditions, which are: The item was randomly assigned with 3 indicators according to the level of difficulty which the system determined to select in the random type selection box as follows: a) Custom randomization can indicate the number of items with difficult, medium, and easy for each of the 3 items in each indicator; b) Randomly selected from the student's year criteria, they can be chosen from years 1-5.

       1.1.3 Examination management consists of a) The process for conducting the test which starts when the test taker presses the button to enter the test. The computer screen shows all 78 items of the test. It is a multiple-choice situational test for the test taker to click on only one of the most correct options. Test takers can change their answers by clicking on the new answer they want after selecting all the answers on each page, click the arrow for next to start the next page, or when

they can't do the item, they can skip that item first and click the arrow for previous to answer that item again. The test takers keep doing the test until they reach the last page by clicking the submit answer button. If the test takers do not select all the answers, the system will show the unselected item and press to select the complete answer and click the resend button. The test time is 1 hour and 30 minutes. The system will automatically display a countdown timer until the timer expires, it will be considered the end of the test or if the test taker completes the test before the specified time, the test can be sent immediately. b) Assessment of the test takers' satisfaction with using the electronic item bank by answering the questionnaire and c)The test report is the part where the test takers can know the test result immediately after the test is completed. The test results are reported as an observed score for each skill and total score, pass/fail assessment results for each skill, and totals in the form of tables and pie charts. The test takers can press to save the test results. When the test results are known, the test is completed.

1.1.4 Evaluation management consists of a) The report of individual examination results showing information on the test sequence, student ID, name, surname, institution, year, grades obtained, time is taken and test scores; b) The report of institutional examination results showing the number of undergraduate students taking the test for each year, the average score of each skill, the average score of each year of each institution; c) The report of the overall examination results of all test takers from all institutions as information about the number of undergraduate students taking the test for each year, the average score for each skill, the average score for each year by showing the comparison results of all institutions that conducted the test; and 4) The results of the questionnaire shows the scores of undergraduate students' satisfaction towards the use of the system, i.e. mean scores and standard deviations.

1.1.5 Scoring management consists of a) managing the test scores of all test takers showing the results of the basic statistical analysis of the test and b) taking the data out of the system to print the test results data of the test takers.

1.2 The quality of the items collected in the electronic item bank was found that there were 279 items, with the items having the discriminant parameters between 0.139 and 2.499. The mean of the discriminant parameters of all items was 0.862 and the difficulty parameters were between -2.827 and +2.688, classified as easy, medium, and difficult items, with the mean of all item's difficulty parameters of 0.069. It was a moderately difficult item and can be classified well as shown in the table.

Table 1: The results of the quality of items based on the multidimensional item response theory (MIRT) for each factor stored in the electronic item bank.

Factors	Discrimination parameters (a)	Difficulty parameters (b) (number of items)			All items
		Easy exams (-3 ≤b<-1)	Medium exams (-1 ≤b<1)	Difficult exams (1 ≤b< 3)	
Problem-solving ability	0.233 - 1.961	-2.309 – (-1.142) (15)	-0.967 – 0.972 (18)	1.110 - 1.912 (17)	50
Intellectual leadership	0.233 – 2.499	-2.031 – (-1.086) (9)	-0.652 – 0.558 (10)	1.013 - 2.027 (9)	28
Interpersonal relationships	0.182 – 2.175	-1.795 - (-1.043) (15)	-0.499 – 0.964 (18)	1.218 - 2.118 (20)	53
Responsibility	0.194 – 2.498	- 2.201 - (-1.139) (7)	-0.907 – 0.765 (7)	1.162 - 2.688 (7)	21
Information technology ability	0.176 – 2.301	- 2.254 - (-1.043) (13)	-0.927 – 0.827 (21)	1.002 - 1.909 (9)	43
Communication ability	0.300 – 2.457	-2.827 - (-1.040) (7)	-0.717 – 0.516 (7)	1.013 - 2.432 (7)	21
Learning management ability	0.174 – 2.135	-1.570 - (-1.126) (9)	-0.829 – 0.395 (15)	1.064 - 1.981 (9)	33
Measurement and evaluation of learner ability	0.139 – 1.853	-1.985 - (-1.104) (9)	-0.665 – 0.383 (12)	1.115 - 2.233 (9)	30
Total	**a average = 0.862**	**b easy average = -1.535 (84)**	**b medium average = 0.082 (108)**	**b difficult average = 1.540 (87)**	**279**

1.3 The quality of the test and the results of the determination of the cut-off scores of the randomized test from the electronic item bank found that:

1.3.1 The results of the structural validity test using multidimensional analysis of the test were analyzed by the competing model. The multidimensional approach, consecutive approach, and composite approach were analyzed based on the coherence of the model with the empirical data by Deviance Statistic (G^2) and Akaike information criterion (AIC) and based on the comparison of harmony between the models by likelihood ratio, which was equal to the difference of G^2 between the models by using the Chi-Square. It can be seen that the test had various statistical values that

showed consistency. The Deviance Statistic (G^2) and Akaike information criterion (AIC) of the multidimensional approach were the lowest compared to the composite approach and consecutive approach. Therefore, it indicated that the model of measuring learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs was the most suitable for the multidimensional model and when testing the difference of G^2 between the composite and multidimensional approach, and consecutive and multidimensional approach was found that there was a statistically significant difference at the .05 level.

Table 2: Structural validity test

Model	G^2	AIC	n
Composite approach	53794.409	53952.409	79
Consecutive approach	54280.224	54452.224	86
Multidimensional approach	53729.526	53917.526	94
Difference of G2 (Model 1 and 3): χ2 = 64.883, df = 15, $p < .05$			
Difference of G2 (Model 2 and 3): χ2 = 550.698, df = 8, $p < .05$			

1.3.2 The results of evaluating item fit according to the multidimensional item response theory (MIRT) with 78 items revealed that OUTFIT MNSQ ranged from .75 to 1.19 and INFIT MNSQ ranged from .85 to 1.15.

1.3.3 The results of the test for the EAP reliability were 0.707. When separated into dimensions, including problem-solving ability, intellectual leadership, interpersonal relationships, responsibility, information technology ability, communication ability, learning management ability, and measurement and evaluation of learner ability, the reliability values were 0.521, 0.548, 0.600, 0.572, 0.531, 0.564, 0.587 and 0.593, respectively.

Table 3: EAP Reliability of the test based on the dimensions

Dimensions	EAP reliability
The whole test (composite approach)	0.707
Problem-solving ability	0.521
intellectual leadership	0.548
interpersonal relationships	0.600
responsibility	0.572
information technology ability	0.531
communication ability	0.564
learning management ability	0.587
measurement and evaluation of learner ability	0.593

1.3.4 The result of determining the cut-off score by analyzing the criterion zone on the Wright map from the ConQuest program. This research had a Wright map showing the criterion

zone of behavioral levels that indicated learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs. Each item corresponded to the test construction and contained 1 threshold, which was a point divided into two parts of learning outcomes between pass and fails. The area above the threshold indicated the competency level that had passed while the area below the threshold indicated the competency level that had failed. From the Wright map, it was found as follows: 1) the area of cognitive skills had the cut-off score of 10 and 6 on the problem-solving ability and intellectual leadership respectively; 2) The area of interpersonal skill and responsibility had the cut-off score of 12 and 3 on the interpersonal relationships and responsibility respectively; 3) The area of numerical analysis, communication, and information technology skills had the cut-off score of 4 and 5 on information technology ability and communication ability respectively; 4) The area of learning management skills had the cut-off score of 3 and 4 on learning management ability and measurement and evaluation of learner ability respectively. This showed that undergraduate students with competence in each area passed the criteria when the score was greater than the cut-off score given both factors as shown in Table 4.

Table 4: The average competency level (the cut-off score) and the cut-off score of each factor.

The 4 areas	8 factors (The total score)	The average competency levels (the cut-off score)	The cut-off score
Cognitive skills (full score: 24 points)	Problem-solving ability (15)	0.284	10
	Intellectual leadership (9)	0.180	6
Interpersonal skill and responsibility (full score: 21 points)	Interpersonal relationships (15)	1.189	12
	Responsibility (6)	0.216	3
Numerical analysis, communication, and information technology skills (full score: 15 points)	Information technology ability (9)	-0.286	4
	Communication ability (6)	1.323	5
learning management skills (full score: 18 points)	Learning management ability (9)	-0.355	3
	Measurement and evaluation of learner ability (9)	-0.212	4
The total		**0.292**	

2. The results of the assessment of the efficiency of the electronic item bank were as follows:

2.1 The electronic item bank was evaluated by 5 experts before being fully released. The result revealed that the electronic item bank was highly rated by experts (M= 4.74, SD= 0.41). As seen in Table 5, the table informed expert rating on each domain; 1) utility (M= 4.70, SD= 0.50), 2) feasibility (M= 4.76, SD= 0.40), 3) propriety (M= 4.76, SD= 0.40), and 4) accuracy (M= 4.73, SD= 0.33).

Table 5: The expert rating on the electronic item bank

No.	Description	M	SD	Rating
1.	*Utility*	*4.70*	*0.50*	*Highest*
1.1	Service was designed around users and their needs.	4.80	0.45	Highest
1.2	Service was responsive, reliable, and worth trying.	4.60	0.55	Highest
1.3	Service was useful for the associated party.	4.60	0.55	Highest
1.4	Information from the service was helpful for substantial development in the educational program.	4.80	0.45	Highest
2.	*Feasibility*	*4.76*	*0.40*	*Highest*
2.1	Service was user-friendly and practical.	5.00	0.00	Highest
2.2	Service was acceptable by general standard.	4.80	0.45	Highest
2.3	The testing outcome was correct and reliable.	4.60	0.55	Highest
2.4	Service was accessible and efficient.	4.80	0.45	Highest
2.5	Service was able to correctly measure the learning outcomes of students based on TQF: HEd.	4.60	0.55	Highest
3.	*Propriety*	*4.76*	*0.40*	*Highest*
3.1	Instruction was informative, clear, and concise.	5.00	0.00	Highest
3.2	Several contents in each testing were appropriate.	4.60	0.55	Highest
3.3	The testing period was noticeable and appropriate.	4.80	0.45	Highest
3.4	The user interface was well designed and workable.	4.60	0.55	Highest
3.5	The testing outcome was applicable.	4.80	0.45	Highest
4.	*Accuracy*	*4.73*	*0.33*	*Highest*
4.1	The testing objective was clear.	5.00	0.00	Highest
4.2	The service algorithm worked precisely.	4.80	0.45	Highest
4.3	The testing report was precise according to the examinee's trait.	4.40	0.55	High
	Total rating	*4.74*	*0.41*	*Highest*

2.2 The results of the undergraduate students' satisfaction assessment for using the electronic item bank revealed that they were highly satisfied (M= 4.64, SD= 0.69). As seen in Table 6, the table informed the rating on each domain; 1) service workability (M= 4.69, SD= 0.61), 2) user experience (M= 4.72, SD= 0.62), 3) user interface: visualization (M= 4.51, SD= 0.84), 4) user interface: design (M= 4.63, SD= 0.71), and 5) user interface: informative and comprehensible (M= 4.63, SD= 0.68).

Table 6: The rating on the electronic item bank

No.	Description	M	SD	Rating
1.	*Service workability*	*4.69*	*0.61*	*Highest*
1.1	Tools on the user interface were workable.	4.72	0.58	Highest
1.2	The service algorithm was appropriate.	4.68	0.60	Highest
1.3	Service was responsive and workable.	4.67	0.64	Highest
2.	*User experience*	*4.72*	*0.62*	*Highest*
2.1	Service responsiveness was acceptable.	4.74	0.60	Highest
2.2	The testing period was noticeable.	4.70	0.67	Highest
2.3	The user got a notification if skipped the item.	4.73	0.60	Highest
3.	*User interface: Visualization*	*4.51*	*0.84*	*Highest*
3.1	The homepage was visually appealing.	4.51	0.85	Highest
3.2	The color tone was well balanced.	4.46	0.86	High
3.3	The user interface design was tidy.	4.57	0.80	Highest
4.	*User interface: Design*	*4.63*	*0.71*	*Highest*

No.	Description	M	SD	Rating
4.1	The filling form was well placed.	4.62	0.71	Highest
4.2	Service tools were well placed.	4.63	0.73	Highest
4.3	Several items per page were appropriate.	4.65	0.70	Highest
5.	*User interface: Informative and comprehensible*	*4.63*	*0.68*	*Highest*
5.1	Text and pictures in the user interface were comprehensible.	4.57	0.75	Highest
5.2	Text and pictures in the user interface were legible.	4.63	0.69	Highest
5.3	The testing report was informative, clear, and concise.	4.70	0.61	Highest
	Total rating	*4.64*	*0.69*	*Highest*

The criteria for interpreting the results were as follows: (Wongrattana, 2017)

The average score 1.00-1.49 was useful/possible/appropriate/correct/lowest satisfaction.

The average score 1.50-2.49 was useful/possible/appropriate/correct/low satisfaction.

The average score 2.50-3.49 was useful/possible/appropriate/correct/moderate satisfaction.

The average score 3.50-4.49 was useful/possible/appropriate/correct/high satisfaction.

The average score 4.50-5.00 was useful/possible/appropriate/correct/highest satisfaction.

Discussion

The results of the research were discussed as follows:

1. The development of the electronic item bank for measuring learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs has a similar process found in a study by Paungsombat and Senarat (2019), developing the item bank for testing and diagnostic programs for problem-solving skills; and Krungkraipetch (2011), developing the item bank and organizing an examination with a computer program. There is a process of creating a test, examining the quality of the test, and selecting a quality item through the criteria collected in the item bank that is made into a database stored on the computer. In addition, the electronic item bank developed by the researcher contains the main components of the computer program following the components of the item bank of Janpla and Wannapiroon (2019). The electronic item bank of this study consists of user management and item management, which helped create items, edit items, and select items by random to arrange test sets according to the different *levels* of difficulty. Therefore, users can select items as needed, which are suitable for use in measuring the skills of students in each year. After completing the management of the item set, users can do both computer-based tests and paper-based tests depending on the user's convenience. If there are no computers, users can print out the tests available in the electronic item bank for paper testing, which is convenient for users. The electronic item bank is managed by logging in to the test via the designated website,

conducting the test, and knowing the results immediately. The system can report the results of individual, institutional, and overall test-takers at all institutions. All test takers' scores are managed and any associated party was able to export needed information from the system. Therefore, it can be seen that the electronic item bank is similar to the development of test item banking on English language substance for Pratomsuksa 4 students by Tangtong (2016), which stores quality items in Google Drive that users can create and edit items, randomly select items based on learning standards and indicators, discrimination parameters and difficulty parameters as well as printing a test that can be selected as needed. Moreover, the number of items available in the item bank for this research is appropriate according to Embretson and Reise (2000). This shows that there is no restriction for several items stored in the electronic item bank, but the items should be item discriminated and difficulty covering all ranges of abilities and there may be at least 100 items. The quality of the items selected by the researcher to be stored in the electronic item bank is consistent with Baker and Kim (2017), criteria of the discrimination parameters, which is the items with the discriminant parameters of 0 or higher selected. The difficulty parameter ranges from -3 to +3 Logic. It can also be seen that the items stored in the item bank are divided into 3 levels of difficulty: 1) Easy Level ($-3 \leq b < -1$) 2), Medium Level ($-1 \leq b < 1$) and 3) Difficult Level ($1 \leq b < 3$). Most of the items are moderate difficulty, but to develop a good item bank, there should be equal or similar items in each difficulty level. However, the development of this item bank is early development. Therefore, the additional item bank can be developed in the future. This is in line with Boonsri, Tangdhanakanond, and Kanjanawasee (2018) for the quality of the learning outcome measurement based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs. Items were randomly acquired from the item bank to form a test to analyze the quality of the test according to the Multidimensional Item Response Theory (MIRT). The results of the analysis show that the items of OUTFIT MNSQ and INFIT MNSQ meet specified criteria and are consistent with Wright and Masters (1982, as cited in Chianchana, 2009) proposing the use of OUTFIT MNSQ and INFIT MNSQ between .75 and 1.33. Lunz et al. (1990) suggest that the OUTFIT MNSQ and INFIT MNSQ ranges between .60 and 1.50, indicating that the model is appropriate for the item. When examining the structural validity test by the competing model among multidimensional approach, consecutive approach, and composite approach, this reveals that the learning outcome measurement model according to the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs is most suitable for the multidimensional approach. The statistical results of the analysis are consistent in the same direction. Therefore, it is sufficient evidence to indicate that the structure of the variables in this research is suitable for the measurement of the multidimensional approach. As for the reliability analysis of the test, it is found that the whole version of the EAP reliability and the reliability value when separated into dimensions. It is consistent with the criteria of Kanjanawasee (2013) who states the criteria

for considering the reliability coefficient of the measuring instruments should not be lower than .50. This indicates that the measurement quality of reliability meets the criteria. To obtain the cut-off score of the measurement, the researcher presents it using the criterion zone on the Wright map. The method for constructing the cut-off score from the criterion zone provides the right interpretation due to analysis according to the Multidimensional Item Response Theory (MIRT).

2. This is an analysis that corresponds to the nature of the multidimensional variables. Thus, it should be helpful for measurement accuracy (Nuansri, 2016) in line with Briggs and Wilson (2003), who explains that if there is statistical evidence to support the method, the measurable trait is multidimensional. As a result, the interpretation of the analysis results is very important. If the measurable traits are multidimensional, then they are analyzed and interpreted as that unidimensional. It is considered an inaccurate, non-representative of the true competent person and will have a significant impact on the process of measurement and assessment that has a high bargaining power or involves decision-making.

3. The importance of evaluating the effectiveness of the electronic item bank for measuring learning outcomes per the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs is to check the system quality of the computer program development. In this research, the electronic item bank is evaluated its efficiency by experts before full release. Using Stufflebeam's evaluation model covers 4 domains: 1) utility, 2) feasibility, 3) propriety, and 4) accuracy. This is like Kanjanawasee (2012), which can assess whether the developed system can be used in practice or not, in line with the guidelines of Chaimongkol (2015). There is an evaluation of the quality of the electronic item bank after actual use with students and participants are asked to evaluate their experience with the electronic item bank. The highlight of this evaluation is to know whether the developed electronic item bank can respond to the use of the sample group or not because the results of the assessment will lead to the development of the electronic item bank in the future. (Chin, Diehl, and Norman, 1988; Navas et al., 2007) Therefore, it can be assured that the electronic item bank is of standard quality and can be used practically, that is, the electronic item bank for measuring learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs is effective.

Implications

1. Educational institutions can use the electronic item bank to measure learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs. In addition, they can also apply to measure cognitive skills, interpersonal skills and responsibility, numerical analysis, communication,

and information technology skills as well as learning management skills among undergraduate students in other fields to use the information obtained from the measurement and evaluation as a guideline for further student development. The implementation should take into account the availability of computer equipment and the internet network because it is an online test that is accessed through a browser such as the Internet Explorer from the given URL. Therefore, the person who uses the test must be proficient in using computers to facilitate the students who take the test using computers. Moreover, during the test teachers should focus on encouraging students to be attentive to the questions and options and try to do the test to the best of their ability.

2. The test developed in the research can be applied to undergraduate students in education programs from year 1 until year 5. The test applied to students in each year should have a different proportion of difficulty and the difficulty level of the test depends on the measurement objectives. For example, if measuring the exit exam, the difficulty level should be increased or if measuring the student development, the proportion of the difficulty level should be reduced. The developed items have difficulty levels including the easy, medium, and hard levels and those who use the tests can specify the difficulty level of the items as needed to obtain information for use in planning, designing learning management, and organizing additional activities for undergraduate students in education programs on an ongoing basis until students achieve the expected learning outcomes.

Recommendations

1. In further studies, the items in the electronic item bank should be developed into a multimedia computerized to encourage the test takers to be more interested during the test or developed into computerized adaptive testing that can be tested to suit the examinee's ability level.

2. In this research, the test is the multiple-choice situational test and the score ranges between 0 and 1. In future research, another model should be developed. Alternatively, a combination of models may be used to measure learning outcomes according to the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs.

3. The development of the electronic item bank based on professional standards and ethics for teachers should be studied or the electronic item bank should be developed that corresponds to the learning outcomes according to the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in other programs.

Conclusion

The electronic exam bank, for measuring learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs, is developed as a source of quality item collection. The quality of the items and the test is examined by applying the Multidimensional Item Response Theory (MIRT). This analytical technique is appropriate for the test structure of the psychological variables so that the validity and reliability of the instruments can provide accurate and useful information. Furthermore, the thing that should be taken into account is the criterion of the test. In this research, after developing the test, the cut-off scores were also set by analyzing the criterion zone on the Wright map from the ConQuest program to be used as the criteria of test to measure learning outcomes based on the Thai Qualifications Framework for Higher Education (TQF: HEd) of undergraduate students in education programs. Furthermore, the electronic item bank is efficient, is suitable, and can be used systematically. Based on user need, the users can select the test and can do both computer-based tests and paper-based tests depending on the user's convenience. In the case of a user who conducts the test using a computer, the test taker can log in to the test system through the designated website, take the test, and know the testing result immediately after testing is done, which makes it convenient and fast. It gives information that reflects the results of the educational management of each institution. The institution is aware of the student teachers' learning outcomes whether they achieve the goals that the curriculum aims for or not. This will lead to the development of students until they become qualified graduate teachers.

Acknowledgments

References

Adams, R. J., Wilson, M. R., & Wang, W.C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1-23.

Angsuchoti, S. (2019). *Evaluation of Sukhothai Thammathirat Open University English Proficiency Test (STOU-EPT) E-testing system*. Nontaburi: Sukhothai Thammathirat Open University.

Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. Springer International Publishing.

Boonsri, S., Tangdhanakanond, K., & Kanjanawasee, S. (2018). Development of Multi-stage Test Item Bank for Measuring Multidimensional Knowledge Competency of Vocational-Technical Teacher Students. *Dusit Thani College Journal, 12*(3), 600-615.

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

Chaimongkol, N. (2015). *System development for on-the-fly assembled multistage adaptive testing with reflective feedback in information technology professional examination* [Unpublished doctoral dissertation]. Chulalongkorn University, Bangkok, Thailand.

Chianchana, C. (2009). Multidimensional analysis. *Journal of Education Khon Kaen University*. *32*(4), 13-22.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). *Development of An Instrument Measuring User Satisfaction of the Human-Computer Interface*. [Paper presentation]. The CHI88: Conference on Human Factors in Computing Systems, Washington D.C., USA.

Comrey, A. L., & Lee, H. B. (2016). *A first course in factor analysis*. Psychology Press.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Inc. Mahwah.

Friedenberg, L. (1995). *Psychological testing: Design, analysis, and use*. Pearson.

Janpla, S. & Wannapiroon, P. (2019). Machine learning for intelligent item banks to enhance the professional competencies of the software developer. *Sripatum Chonburi Journal*, *15*(3), 183–196.

Kanjanawasee, S. (2012). *Modern test Theories*. Bangkok: Chulalongkorn University Press.

Kanjanawasee, S. (2013). *Classical test Theories*. Bangkok: Chulalongkorn University Press.

Krungkraipetch, N. (2011). Development of computerized item bank and online examination. *The Journal of Faculty of Nursing Burapha University*, *19*(2). 1-13.

Lunz, M., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 331-345.

Navas, H., Osornio, A. L., Baum, A., Gomez, A., Luna, D., & Quiros, F. G. B. d. (2007). *Creation and Evaluation of a Terminology Server for the Interactive Coding of Discharge Summaries*. [Paper presentation]. The MEDINFO 2007, Brisbane, Australia.

Nuansri, M. (2016). *Development and investigation of multidimensional construct of responsible citizenship of lower secondary school students: an application of construct mapping approach* [Unpublished doctoral dissertation]. Chulalongkorn University, Bangkok, Thailand.

Office of the Higher Education Commission. (2009). *Thai Qualifications Framework for Higher Education*. http://www.mua.go.th/users/tqf-hed/

Office of the Higher Education Commission. (2020). Higher education statistics: Academic year of 2019. http://www.info.mua.go.th/info/

Paungsombat, K., Senarat, S., & Senarat, B. (2019). Development of an item bank for testing and diagnostic of problem-solving skills. *Journal of Graduate School, 16*(74), 27-39.

Reckase, M. D. (2009). *Multidimensional item response theory*. Springer-Verlag New York.

Rohitasthira, B. (2016). *National policy for teacher production in 21st century*. https://www.moe.go.th/websm/2016/sep/407.html

Tangtong, S. (2016). Development of Test Item Banking on English Learning Substance For Pratomsuksa 4 Students. *Journal of Information*, *15*(1), 81-89.

Vollakitkasemskul, S. (2011). *Applied statistics for behavioral and social science research*. Education Faculty Udon Thani Rajabhat University.

Wongrattana, C. (2017). S*tatistics for research*. Chulalongkorn University Book Center.