

The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates

Satien Janpla^{*}, Pallop Piriyastrawong

Division of Information and Communication Technology for Education, King Mongkut's University of Technology North Bangkok, Thailand

Received June 5, 2020; Revised July 9, 2020; Accepted August 10, 2020

Cite This Paper in the following Citation Styles

(a): [1] Satien Janpla, Pallop Piriyastrawong, "The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates," *Universal Journal of Educational Research*, Vol. 8, No. 9, pp. 4163-4172, 2020. DOI: 10.13189/ujer.2020.080942.

(b): Satien Janpla, Pallop Piriyastrawong (2020). *The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates*. *Universal Journal of Educational Research*, 8(9), 4163-4172. DOI: 10.13189/ujer.2020.080942.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The objectives of this research were 1) to develop an Intelligent Multilevel Item Bank (I-MIB) Model for the National Evaluation of Undergraduates and 2) to assess the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates. The results of the research were as follows: 1. The I-MIB Model for the National Evaluation of Undergraduates consisted of four parts: Input, Process, Output, Feedback. The Input consisted of 1. Curriculum, 2. Subject, 3. Topic, 4. Objectives, 5. Cognitive Skills, 6. Teacher, 7. Student, and 8. Program Designated Lecturers. The Process consisted of 1. Question Management, 2. Cognitive Skills Processing, 3. Data Mining, 4. Trial Testing, 5. Improving Questions, 6. Testing, 7. Proficiency Classification, and 8. Suggestions. The Output consisted of 1. Testing Report, 2. Proficiency Level, and 3. Suggestions. Finally, the Feedback consisted of returning the evaluation results to improve the input and process and yield accurate evaluation results. 2. The assessment of the appropriateness of the I-MIB model for the national evaluation of undergraduates included 13 experts with an expertise in ICT for education, information technology, computer science, and educational measurement and evaluation. The evaluation results found that the I-MIB model for the evaluation of national undergraduates was suitable at the highest level with a mean of 4.60 and a standard deviation of 0.58.

Keywords Intelligent Multilevel Item Bank, Classical Test Theory, Data Mining, Revised Bloom's Taxonomy

1. Introduction

Thailand has established a system for educational quality assurance under the National Education Act of 1999 and its amendment (Version 2) in 2002, which supports the development of educational quality and standards at all levels. The Office of the Higher Education Commission has established a framework for higher education qualifications and standards and announced the National Quality Standards for Higher Education in 2009 with the objectives of meeting higher education standards and guaranteeing the quality of graduates at each qualification level and in all subject areas. Higher education institutions must use a standardized qualification framework as a guideline for the development or improvement of curriculum and teaching to ensure that graduates produced at the same qualification level are held to comparable quality standards, both at the national level and international level [1].

Establishing educational measurement and evaluation standards is very important for teaching and learning because they are indicators that graduates had learning characteristics or results that meet the standard higher education qualifications [2] and are the result of knowledge development from the teaching and learning process of each institution. Academic achievement is, therefore, a measure of the quality of the performance of the important responsibilities of the school. [3] One type of educational

measurement tool that is widely used and works well is the item, and the most commonly used item is the multiple-choice test. This is because the test can measure every level of achievement according to Bloom's educational objective classification, from the knowledge level to the evaluation level. Therefore, most instructors prefer to use multiple-choice tests to measure student achievement. [4]

Currently, one problem of using this form of test is that there are still many instructors who use the test once because there is no systematic, safe, and convenient storage method. [4] Achievement tests and multiple-choice tests are often of poor quality and are not standardized because the teacher who produces the tests lacks knowledge of good test building skills, the tests are inconsistent with the curriculum, the tests do not measure the appropriate indicators, the tests do not correspond to the learning theory of Bloom, and neither type of test is typically assessed for quality [3]. The item analysis process is complicated and requires a great deal of time to conduct. [5]. Besides, another problem is that there is no central item that can be used to test every university that is evaluated based on qualification frameworks, to ensure the quality of the graduate education at each level and in each subject.

Creating an item bank is one way to solve this problem. The present research concerns the Intelligent Multilevel Item Bank (I-MIB) Model for the National Evaluation of Undergraduates system, which is a system that will help manage items so that they can be easily found and used to systematically, appropriately and conveniently improve the quality of tests to better measure learning and teaching. The item bank must be built according to the learning objectives and must be an automated system that can analyze a test based on information obtained from the answers of the test takers to produce tests with different levels of difficulty following the standards. Most importantly, Bloom's cognitive level of each test item is determined, and the system will automatically process the appropriate Bloom's cognitive level using data mining techniques, thereby aiding the instructor. There is no need to assign a Bloom's cognitive level to the test itself, which can be a challenge for some instructors since the items selected can generate tests according to the proficiency level of the test takers and the level desired by the instructors, such as Foundation Level, Advanced Level, and Expert Level. The system can also advise candidates by using data mining techniques to assist in processing and finding relationships with data from other examinations to provide recommendations to test takers about their weaknesses, strengths, or objectives. Besides, the test results can be used as guidelines for the development or improvement of the curriculum, teaching and learning to ensure that graduates who achieved the same qualification level meet comparable quality standards, at both national and international levels.

2. Objectives

1. To develop the I-MIB Model for the National Evaluation of Undergraduates.
2. To assess the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates.

3. Literature Review

3.1. Intelligent Item Bank

An Intelligent Item Bank System is a set of examination items built to the same standards by using statistical data from the tests of test-takers, thus making the test inventory system intelligent. It can be used as an assessment tool. Items may be selected from the test storage system according to difficulty level and standards. Once the item is completed, the instructor can assess the knowledge of the candidate and could advise them on ways to improve if they are below the standard [6], [7].

3.2. Classical Test Theory

Charles Spearman proposed a traditional test model, which used general knowledge of how to check the relationship between observable scores and true scores [8], for item quality analysis. The quality of the test created for evaluation of the student's learning depended on the test's planning and construction, the depth of the subject knowledge of the item author, and the author's test writing skills. An item that was created must be tested, or the test results can be used to analyze the quality of the test. [9]. Item analysis is an examination technique for item quality by item. According to an important traditional theory, item quality analysis is based on item difficulty. An estimate of the difficulty can be obtained using the proportion of people who answered correctly, which is calculated as the number of respondents who answered correctly divided by the total number of respondents. Item discrimination meant that the test could separate people with knowledge from those who did not have measurable knowledge. The analysis results determine whether each item has performed its function properly or not [10], [11], [12].

3.3. Revised Bloom's Taxonomy

The taxonomy of Benjamin S. Bloom was published in the year 1956. It classified human learning into 3 areas: Cognitive, Affective, and Psychomotor. Abilities in each area were classified from lowest to highest. The cognitive domain included knowledge, comprehension, application, analysis, synthesis, and evaluation [13]. In 2001, Bloom's Taxonomy was revised by Anderson & Krathwohl. The cognitive process emphasized the student's learning process in achieving the desired outcomes. Teachers used appropriate cognitive processes and types of knowledge to adjust curriculums and teaching styles to enhance learners'

potential. Bloom's taxonomy has been the most used tool to evaluate academic goals such as general education items or entrance items at all levels, and written tests were used to evaluate the cognitive abilities of students [14]. In the taxonomy, cognitive abilities were categorized as Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating [13], [15].

3.4. Data Mining

Data mining is the process of finding relationships and constructing models by analyzing large amounts of data using machine learning methods, and searching for knowledge from large amounts of data or large databases to extract knowledge or important information that was hidden in the data for analysis or prediction. There are many data mining techniques used to find interesting models in big data, such as clustering, classification, estimation, association rules, prediction, segmentation, time series, anomaly detection, visualization, and others. Furthermore, the methods are interdisciplinary and include applications in fields such as Astronomy, Business, Computer Science, Economics, and others [16], [17], [18].

4. Methodology

This research was a documentary study to study documents and research related to the components of the Item Banks system, by analyzing content analysis to be the starting point for the development of an intelligent multilevel item bank model for the national evaluation of undergraduates. Therefore, this process consisted of a document study, the development of an intelligent multilevel item bank model for the national evaluation of undergraduates. And the suitability assessment of an intelligent multilevel item bank model for the national evaluation of undergraduates developed, which had the following details:

Step 1: Document Study

Methods of document study, the researcher has studied information, consisting of documents, textbooks, researches, articles, dissertations, and a doctoral thesis on the issues regarding the process of creating an item bank to summarize the steps that have been summarized into principles steps for creating an item bank. Documents used in the study, the researcher studied the document by selecting specific criteria and criteria for selecting the document used in the study, it was a document or article with content about creating Standard Examination, item Analysis, item bank Procedure, Classical Test Theory, Revised Bloom's Taxonomy, Multilevel, Data Mining. When choosing a document, verify the information by considering the credibility, considering the source of the document, the year of publication, the location of the

publication, and the name of the publisher that was complete, reliable, or not.

Step 2: Model development

Based on document studies and research, the researcher has designed an intelligent multilevel item bank model for the national evaluation of undergraduates following the Universal Systems Model, consisting of four basic components: input, process, output, and feedback.

Step 3: Assessment of suitability

The tools used in evaluating the suitability of an intelligent multilevel item bank model for the national evaluation of undergraduates were the questionnaire. Next, the expert selection process. Selection with a specific method by 13 experts, which the experts must have specialization in ICT for education, Information Technology, Computer Science, Educational Measurement, and Evaluation. After that, an intelligent multilevel item bank model for the national evaluation of undergraduates would be evaluated by experts for the suitability of the model.

5. Results

5.1. The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates

When developing the model, the researchers designed it according to the concept of the Universal Systems Model [19], [20]. The Universal Systems Model is a general concept of how to represent a process. There are 4 basic components, i.e., input, process, output, and feedback.

Based on the theoretical framework and model design guidelines of the Universal Systems Model, the researchers developed the I-MIB Model for the National Evaluation of Undergraduates, consisting of four parts as follows. Input consisted of 1. Curriculum, 2. Subject, 3. Topic, 4. Objectives, 5. Cognitive Skills, 6. Teacher, 7. Student, and 8. Program Designated Lecturers. The Process consisted of 1. Question Management, 2. Cognitive Skills Processing, 3. Data Mining, 4. Trial Testing, 5. Improving Questions, 6. Testing, 7. Proficiency Classification, and 8. Suggestions. The output consisted of 1. Testing Reports, 2. Proficiency Levels, and 3. Suggestions. Finally, Feedback consisted of returning the evaluation results to improve the input and process and yield accurate evaluation results. These parts are shown in Figure 1.

Model description

The model consisted of 4 parts, namely, input, process, output, and feedback, with the following details.

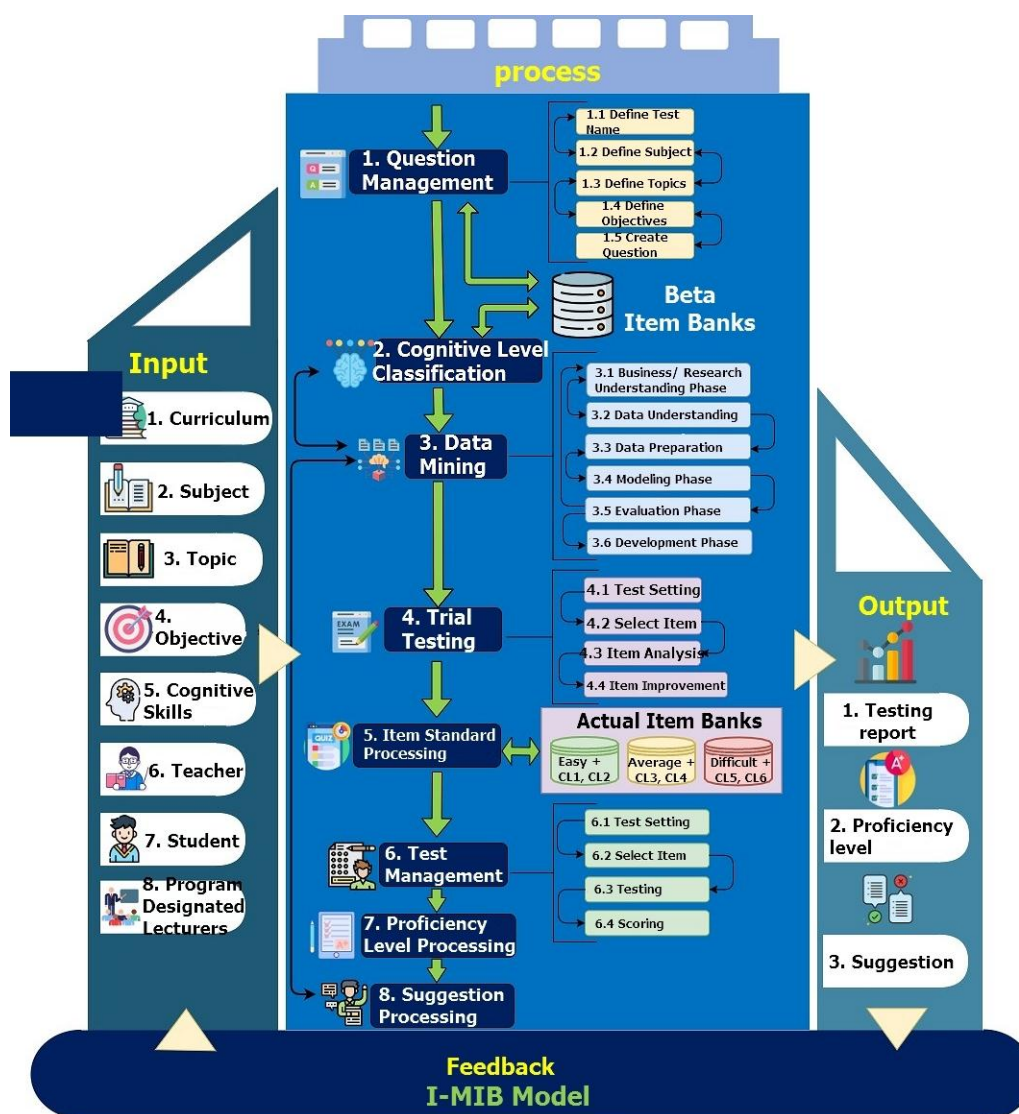


Figure 1. I-MIB model for the national evaluation of undergraduates

Input

1. Curriculum

Curriculum refers to an academic bachelor's degree program that aims to produce graduates with both theoretical and practical knowledge and emphasizes academic knowledge and skills. According to the framework of the national higher education qualifications and standards, when the students completed the entire curriculum at the specified time, they were considered to have completed the course.

2. Subject

The subject was the course offered in the curriculum that will be the source of the bank of items. It defines the scope of what the learner must study in the entire content of the curriculum and is categorized as the objectives of the course that the learners must meet. The content related to the subject in each learning objective must be arranged according to the most suitable steps so that it is easy for learners to understand. Various measurement and

evaluation tools were designed.

3. Topic

The Topic was the subheading of the course corresponding to the test.

4. Objectives

Objectives were messages that specified the learning characteristics and abilities that teachers wanted the students to learn after the students had gone through the teaching and learning activities in a specific subject or chapter. The objective was the medium to create understanding between teachers and learners to have the same purpose of teaching and learning.

5. Cognitive Skills

Cognitive Skills reflected the level of the item according to the Cognitive Skills of the Revised Bloom's Taxonomy. They could differentiate between "High-level thinking and low-level thinking" and in order from basic to complex. They have been divided into six stages,

namely, remembering, understanding, applying, analyzing, evaluating, and creating.

6. Teacher

The teacher was a full-time teacher who had direct qualifications related to the program and was responsible for teaching and research in that field. The teacher had qualifications that were relevant to the program subject and was required to have previously taught or had expertise in the subject to take the test.

7. Student

The student was studying at a bachelor's degree level at a tertiary institution and has completed the course that was being tested.

8. Program Designated Lecturers

Program Designated Lecturers were course professors with an obligation to manage and develop curriculum and instruction, including planning, quality control, follow-up, evaluation, and course development.

Process

1. Question Management

The question management process consisted of the following steps.

1) Define Test Name

This defined the name of the item.

2) Define Subject

This defined the subject of the test that will be made.

3) Define Topics

Define Topics set the topic or lesson set.

4) Define Objectives

Defining objectives set what the learners needed to know and what they needed to be able to do in the form of observable behavior. The learning objectives were derived from the curriculum levels, the curriculum objectives, the purpose of the program, the professional standards, the field of study, and the work up to the course level, including the course objectives, course standards, and descriptions. The desired learning outcomes are used to achieve the objectives of the course.

5) Create Question

This step created questions based on the subjects and objectives according to the number of items planned.

2. Cognitive Level Classification

Each of these items could be classified according to Bloom's taxonomy, but many test-takers or instructors might not classify them correctly; therefore, data mining techniques should be used to help. The process to classify the cognitive level had the following steps: 1. Data Collection, 2. Feature Extraction, 3. Word segmentation, 4. Indexing, 5. Model, and 6. Model Performance

Assessment. The test methods compare the efficiency of the model with the efficiency of the cognitive level of the test by determining the accuracy of the cognitive level of the test [21], [22], [23], [24].

3. Data Mining

Data mining is a process that involves using large amounts of data to find models, approaches, and relationships hidden in that data set; and its use is based on statistical, recognition, machine learning, and mathematical principles. Search rules are used for management, classification, and prediction in decision support processes [25]. Data mining is a step-by-step procedure to make the process efficient and accurate. The data mining procedure has been standardized, and this standard was called the Cross-Industry Standard Process for Data Mining (CRISP-DM) [26], [27], [28]. There were six steps in this standard as follows.

1) Business/Research Understanding Phase

This step is the process that determines the goals and objectives of the data mining process. There must be a plan for the implementation of the process to achieve the goals and objectives.

2) Data Understanding Phase

This is the process of collecting different data for analysis, including studying the features and characteristics of the data, evaluating the quality of the data, and selecting the data to be analyzed. If necessary, after completing this step, additional goals may be found for the analysis. In this case, the researcher returns to step 1 to revise the goals and objectives of the new data mining process.

3) Data Preparation Phase

The goals of this step are to prepare the data, select the data samples, and determine the variables to be analyzed. Additionally, data may be changed to a form that can be analyzed, including eliminating unusual data that may cause the analysis to be inaccurate.

4) Modeling Phase

The goals in this step are to select the data analysis method and then analyze the data to obtain the best results. If necessary, the researcher can return to step 3 to prepare and select samples of additional data and/or add variables to be analyzed.

5) Evaluation Phase

This is the process for evaluating the data mining results obtained from step 4 and determining whether the results can be used for the project. This step may return to the previous step to get more complete results.

6) Deployment Phase

This is the final step after obtaining the complete results from step 5 that will be put into actual use. The results from its use in real work can be evaluated and used for

planning future data mining projects.

Data mining based on CRISP-DM is cyclic and can be adaptive. The operation can return to the previous step to improve its operation and obtain better results before continuing with the next steps.

4. Trial Testing

1) Test Setting

The test set includes setting the test name, item time, date, time, and password.

2) Select Item

The item selection step involves selecting the item from the temporary item bank and bringing it into the standardized examination item bank. The same number of topics are selected in each test to analyze all the questions.

3) Item Analysis

Item analysis was conducted according to Classical Test Theory. The analysis included the item difficulty (p) and classification power (r).

4) Item Improvement

An item that did not pass the standardized requirements must be improved, for example, by adjusting problems or adjusting options, then returned to the test, and finally analyzed in the context of the test again.

5. Item Standard Processing

The standardized examination process will include items with the appropriate difficulty level according to the division of the levels as follows: easy level items, with p values from 0.60-0.79, average with p values from 0.40-0.59, and difficult with p values from 0.20-0.39. The cognitive levels are divided into 3 levels as follows: The Easy Cognitive Level included Remembering and Understanding, the Average Cognitive Level included Applying and Analyzing, and the Difficult Cognitive Level included Evaluating and Creating. Besides, the test difficulty was divided into 3 levels: Foundation, Advanced, and Expert. The details are given in Table 1.

6. Test Management

1) Test Setting

The Test Setting step sets the test name, item time, date, time, and password for the item.

2) Select Item

The select item step will select an item from the item bank that meets the indicated item level, item standard, and specified criteria.

3) Testing

During testing, only those who have registered in advance will be eligible to take the item. The candidates must use their username and password to log into the item. Then, when logged in, they click on the item that they want to take. The item will be displayed. Then, the candidates complete all the questions and then submit their answers. While taking the item, they can see the remaining time for the item. When the time is up, the system terminates the item immediately.

4) Scoring

The system will process the test scores of the candidates as follows.

1. Individual test scores: The system will report the test scores of each candidate and inform the instructor of the individual examination results.
2. Curriculum scores: The system will report the item scores for the curriculum scores; sort the test scores in descending order; and give the highest, lowest, and average scores.
3. University scores: The system will create a score report for each university, which includes the scores for that university.

7. Proficiency Level Processing

This process finds the proficiency levels of students who take the item by using the grade level; and the number of levels that an instructor wants to divide the scores into can be set, such as 3 levels, 4 levels, 5 levels, or 6 levels.

8. Suggestion Processing

It processes the test scores for each topic for each candidate to determine the relationship of any events that occur simultaneously or not, such as if the candidate passed both topic 1 and topic 2, or, on the contrary, if the candidate does not pass topic 3 but does pass topic 4. This type of processing uses the Association Rule. For example, the system will advise the candidate that they go back to study a specific topic if the candidate does not pass the item.

Table 1. Item levels

Test Level	Item Difficulty	Cognitive Levels	% of items that are Easy, Average, and Difficult, respectively
Foundation Level	Easy P 0.60- 0.79	Remembering Understanding	50%, 25%, and 25%
Advanced Level	Average P 0.40 – 0.59	Applying Analyzing	25%, 50%, and 25%
Expert Level	Difficult P 0.20 – 0.39	Evaluating Creating	25%, 25%, and 50%

Output

1. Testing Report

The testing report includes the following:

1. Report on the individual item scores,
2. Report on the Curriculum scores, and
3. Report on the University scores.

2. Proficiency Level

There are 6 proficiency levels of students who take the item: 1. Basic, 2. Intermediate, 3. Advanced, 4. Expert, 5. Excellent, and 6. Proficient.

3. Suggestions

The suggestions from item results will report the topics that were passed and topics that were not passed and will suggest which topics a student should study to pass the item.

Feedback

The results from the output (Testing Report, Proficiency Level, and Suggestions) are analyzed and the results are used to improve the Input and Process. This includes the instructor checking the examination report to see the test scores, seeing which topics have high scores, and seeing which topics have low scores. The results will be used by those responsible for the course to improve the teaching and learning of subjects with low scores and improve the curriculum.

5.2. The Results for the Appropriateness of the Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates

The analysis of the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates used descriptive statistics such as the mean and standard deviation. The evaluation criterion for the model's

appropriateness was a rating scale that used a 5 point Likert Scale, which was the criterion for determining the weight of the assessment. [29]

The interpretation and scoring criteria used to classify the average score of the appropriateness of experts were as follows:

A mean of 4.50-5.00	means	the most appropriate,
A mean of 3.50-4.49	means	very appropriate,
A mean of 2.50-3.49	means	moderate
	appropriateness,	
A mean of 1.50-2.49	means	less appropriate, and
A mean of 1.00-1.49	means	the least appropriate.

The evaluation used 13 ICT experts from education, information, information technology, computer science, and educational measurement and evaluation. The evaluation results are shown in Table 2.

The results of the evaluation of the appropriateness of 13 experts on the I-MIB Model for the National Evaluation of Undergraduates agreed that the process was the most appropriate with a total average of 4.66 and a standard deviation of 0.48. This was followed by feedback being the second most appropriate with a total average of 4.62 and a standard deviation of 0.49. The next were inputs and results. Both were tied for the third most appropriate with total averages of 4.56 and the standard deviations were 0.67 and 0.68, respectively. Finally, the total appropriateness reached most appropriate with an average of 4.60 and a standard deviation of 0.58, which was in the range 0 to 1, and thus it could be considered reliable data. Therefore, it could be concluded that the I-MIB Model for the National Evaluation of Undergraduates has been well prepared, and experts have evaluated it and agreed that it was highly appropriate.

For evaluating some items with high SD values because some experts thought that the list was not appropriate, such as the input section in Teacher SD 0.83 Student SD 0.77 and Program Designated Lecturers SD 0.95, etc.

Table 2. The evaluation results of the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates

Details of the Models	Evaluation results		
	(\bar{X})	(S.D.)	Level of appropriateness
Input			
1. Curriculum	4.77	0.44	Most
2. Subject	4.69	0.48	Most
3. Topics	4.85	0.38	Most
4. Objectives	4.77	0.44	Most
5. Cognitive Skills	4.69	0.48	Most
6. Teacher	4.23	0.83	Very
7. Student	4.38	0.77	Very
8. Program Designated Lecturers	4.08	0.95	Very
Total average	4.56	0.67	Most
Process			
1. Question Management	4.62	0.51	Most
1.1 Define Test Name	4.77	0.44	Most
1.2 Define Subject	4.85	0.38	Most
1.3 Define Topics	4.85	0.38	Most
1.4 Define Objectives	4.62	0.65	Most
1.5 Create Questions	4.69	0.48	Most
2. Cognitive Level Classification	4.62	0.51	Most
3. Data Mining	4.77	0.44	Most
3.1 Business/Research Understanding Phase	4.62	0.51	Most
3.2 Data Understanding Phase	4.62	0.51	Most
3.3 Data Preparation Phase	4.62	0.51	Most
3.4 Modeling Phase	4.69	0.48	Most
3.5 Evaluation Phase	4.69	0.48	Most
3.6 Deployment Phase	4.62	0.51	Most
4. Trial Testing	4.62	0.51	Most
4.1 Test Setting	4.77	0.44	Most
4.2 Select Items	4.62	0.51	Most
4.3 Item Analysis	4.62	0.51	Most
4.4 Item Improvement	4.62	0.51	Most
5. Item Standard Processing	4.62	0.51	Most
6. Test Management	4.62	0.51	Most
6.1 Test Setting	4.54	0.52	Most
6.2 Select Items	4.62	0.51	Most
6.3 Testing	4.62	0.51	Most
6.4 Scoring	4.77	0.44	Most
7. Proficiency Level Processing	4.85	0.38	Most
8. Suggestion Processing	4.85	0.38	Most
Total average	4.66	0.48	Most
Output			
1. Testing Report	4.69	0.63	Most
2. Proficiency Level	4.54	0.78	Most
3. Suggestions	4.46	0.66	Very
Total average	4.56	0.68	Most
Feedback			
1. The item analysis results	4.69	0.48	Most
2. The proficiency level analysis results	4.69	0.48	Most
3. The suggestion analysis results	4.46	0.52	Very
Average	4.62	0.49	Most
Total average	4.60	0.58	Most

6. Discussion

This research on the evaluation of the appropriateness of the I-MIB system for the National Evaluation of Undergraduates creates an examination system consisting of a computer-based test system (CBT) and an online test (OLT) designed using the theory of data mining and Bloom's revised taxonomy. This is in line with the research [30] that studied a computer-based test system by using test selection strategies, data mining, genetic algorithms, and Bloom's revised taxonomy, which ensures high-quality testing. It is also following research [31] that designed a system for self-learning media using the Design of Self-directed e-Learning Material Recommendation System with On-line Evaluation. Since classical test theory (CTT) has been used to create higher education items, which are considered important tools for improving the quality of education and assessment, the item bank helps to manage computer learning achievement tests and can flexibly be used for both summative purposes and formative purposes. The use of traditional theories in the item bank system creates a balance between costs and benefits, which is consistent with research [32] that studied a cost-benefit analysis for developing item banks in higher education.

The evaluation of the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates is performed for all four aspects. The evaluation results are as follows. Input is at the most appropriate level (\bar{X} 4.56, S.D. 0.67) as is the process (\bar{X} 4.66, S.D. 0.48), output (\bar{X} 4.56, S.D. 0.68) and Feedback (\bar{X} 4.62, S.D. 0.49). Finally, all four aspects in total reach the most appropriate level (\bar{X} 4.60, S.D. 0.58). Regarding the input, there will be three types of relevant persons: The Teacher, the Student, and Program Designated Lecturers. This follows the research [32] that analyzed and designed an online item bank system with an automatic item analysis cycle. When there are only two people involved in the system, the teacher and student, the part of the process consisting of Question Management, Item Standard Processing, and test management, following [32] and [33], is at the most appropriate level in every step (\bar{X} 4.66, S.D. 0.48). The Cognitive Level Classification processes use data mining techniques to automatically group the items according to Bloom's taxonomy, which is consistent with the research of [28], [29], [30], and [34] that grouped and classified questions based on Bloom's Revised Taxonomy.

7. Conclusions

The research methodology is divided into 2 steps, which are the following: Step 1: The Development of the I-MIB Model for the National Evaluation of Undergraduates, and Step 2: Assessing the appropriateness of the I-MIB Model for the National Evaluation of Undergraduates.

The results of the research are as follows.

1. The I-MIB Model for the National Evaluation of Undergraduates consists of 4 parts. Input: consists of 1. Curriculum, 2. Subject, 3. Topics, 4. Objectives, 5. Cognitive Skills, 6. Teacher, 7. Student, and 8. Program Designated Lecturers. Process: consists of 1. Question Management, 2. Cognitive Skills Processing, 3. Data Mining, 4. Trial Testing, 5. Improving Questions, 6. Testing, 7. Proficiency Classification, and 8. Suggestions. Output: consists of 1. Testing Report, 2. Proficiency Level, and 3. Suggestions. Feedback consists of 1. Item analysis results, 2. Proficiency level analysis results, and 3. Suggestion analysis results.
2. The appropriateness of the I-MIB Model for the National Evaluation of Under-graduates was evaluated by 13 experts, and the statistics used in the research are the mean and standard deviation. The evaluation results found that the I-MIB Model for the National Evaluation of Undergraduates was most appropriate with a total average of 4.60 and a standard deviation of 0.58.

Acknowledgments

I would like to thank the Graduate Collage of King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand for providing financial support for the research fund.

REFERENCES

- [1] Ministry of Education, "Announcement of the Ministry of Education on the National Qualifications Standards for Higher Education 2009." pp. 1-3, 2009.
- [2] R. Chaiming, "Development of Tools Measuring and Evaluating Learning Outcomes based on Thai Qualification Framework for Higher Education, for the Bachelor Program of Public Health (Community Public Health), Colleges under Praboromarajchanok Institute," *J. Nurs. Educ.*, vol. 7, no. 4, pp. 168-183, 2014.
- [3] T. Panprasert, "The Development of Thai Language Achievement Test for Grade 6 Accordance to Core Basic Education Curriculum B.E. 2551," *Veridian E-Journal*, vol. 7, no. 1, pp. 261-279, 2014.
- [4] S. Tangtong, "Development of Test Item Banking on English Learning Substance for Pratomsuksa 4 Students," *J. Inf.*, vol. 15, no. 1, pp. 81-89, 2016.
- [5] N. Nakaew, "A Development of Management System for Examination Bank and Test on Internet," in *The 5th Rajamangala University of Technology National Conference*, 2013, pp. 48-57.
- [6] S. Janpla and P. Wanapiron, "System framework for an intelligent question bank and examination system," *Int. J.*

- Mach. Learn. Comput.*, vol. 8, no. 5, pp. 488–494, 2018.
- [7] S. Janpla and P. Wannapiroon, “Machine Learning for Intelligent Item bank to Enhance the Professional Competencies of Software Developer,” *Sripatum Chonburi J.*, vol. 15, no. 3, pp. 183–196, 2019.
- [8] Y. He, “Analysis of a self-learning system of English flipped classroom based on adaptive algorithm,” *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 8, pp. 103–116, 2018.
- [9] S. Kanjanawasee, *Classical Test Theory*, 4th ed. Bangkok: Chulalongkorn University Printing House, 2001.
- [10] S. Muangkhoua, “Test Analysis with Classical Test Theory (CTT): Finding Item Difficulty and Item Discrimination using SPSS Program,” *Vajira Med. J. J. Urban Med.*, vol. 61, no. 4, pp. 301–318, 2017.
- [11] I. Himelfarb, “A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating,” *J. Chiropr. Educ.*, vol. 33, no. 2, pp. 151–163, 2019.
- [12] J. Liu, B. Dong, W. Zhou, and Z. Qian, “Optimal selection of hydraulic indexes with classical test theory to compare hydraulic performance of constructed wetlands,” *Ecol. Eng.*, vol. 143, no. December, p. 105687, 2020.
- [13] R. Pradubwate, “Technology Learning Management Guidelines by using Bloom’s taxonomy,” *Veridian E-Journal, Silpakorn Univ.*, vol. 10, no. 3, pp. 1051–1065, 2017.
- [14] Y. H. Chang and H. W. Chen, “An automatic inference system for the quality analysis of test items based on the bloom’s revised taxonomy,” *Proc. 2009 Int. Conf. Mach. Learn. Cybern.*, vol. 5, no. July, pp. 2852–2856, 2009.
- [15] P. N. Sagala and A. Andriani, “Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom’s Taxonomy,” *J. Phys. Conf. Ser.*, vol. 1188, no. 1, 2019.
- [16] M. Goyal and R. Vohra, “Applications of Data Mining in Higher Education,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 113–120, 2012.
- [17] N. Ketui, W. Wisomka, and K. Homjun, “Using classification data mining techniques for students performance prediction,” *ECTI DAMT-NCON 2019 - 4th Int. Conf. Digit. Arts, Media Technol. 2nd ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng.*, pp. 359–363, 2019.
- [18] M. Khanbabaei, M. Alborzi, F. M. Sobhani, and R. Radfar, “Applying clustering and classification data mining techniques for competitive and knowledge-intensive processes improvement,” *Knowl. Process Manag.*, vol. 26, no. 2, pp. 123–139, 2019.
- [19] J. Flowers, “The problem in technology education (A definite article),” *J. Technol. Educ.*, vol. 21, no. 2, pp. 10–20, 2010.
- [20] N. Sardone, “Building Bots to Develop Systems Thinking,” *Sci. Scope*, vol. 040, no. 05, pp. 32–42, 2017.
- [21] S. Janpla, “The effectiveness of automated Thai documents categorization based on machine learning,” *J. Theor. Appl. Inf. Technol.*, vol. 66, no. 1, pp. 43–47, 2014.
- [22] D. A. Abduljabbar and N. Omar, “Exam questions classification based on Bloom’s taxonomy cognitive level using classifiers combination,” *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 447–455, 2015.
- [23] A. Osman and A. A. Yahya, “Classifications of Exam Questions Using Linguistically- Motivated Features: A Case Study Based on Bloom’s Taxonomy,” *Sixth Int. Arab Conf. Qual. Assur. High. Educ.*, vol. 2016, no. February 2016.
- [24] S. K. Patil and M. M. Shreyas, “A Comparative Study of Question Bank Classification based on Revised Bloom’s Taxonomy using SVM and K-NN,” *2017 2nd Int. Conf. Emerg. Comput. Inf. Technol. ICECIT 2017*, pp. 1–7, 2018.
- [25] M. El Mabrouk, S. Gaou, and M. K. Rtili, “Towards an intelligent hybrid recommendation system for E-learning platforms using data mining,” *Int. J. Emerg. Technol. Learn.*, vol. 12, no. 6, pp. 52–76, 2017.
- [26] A. Mungekar and N. Parab, “Quora Insincere Questions Classification,” no. July 2019.
- [27] F. Mart, L. Contreras-ochando, M. Kull, and N. Lachiche, “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories,” vol. 4347, no. c, 2019.
- [28] U. Shafique and H. Qaiser, “A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA),” *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014.
- [29] C. Paiwithayasiritham, P. Makmee, and K. Mingsiritham, “The Development of a training program for basic education school teachers on developing the higher level learning assessment methods,” *Veridian E-Journal, Silpakorn Univ.*, vol. 6, no. 1, pp. 904–913, 2013.
- [30] M. H. Ying, S. H. Huang, and L. R. Wu, “An item selection strategy based on association rules and genetic algorithms,” *2009 4th Int. Conf. Innov. Comput. Inf. Control. ICICIC 2009*, pp. 1040–1044, 2009.
- [31] F. J. Liu, “Design of self-directed e-learning material recommendation system with on-line evaluation,” *Proc. - 2008 Int. Conf. Converg. Hybrid Inf. Technol. ICHIT 2008*, pp. 274–278, 2008.
- [32] S. Draaijer, *A cost-benefit analysis for developing item banks in higher education*, vol. 1014. Springer International Publishing, 2019.
- [33] S. Damrong and B. Pudsadee, “An Online Item Banking System Using Item Analysis,” *Eighth Natl. Conf. Comput. Inf. Technol.*, pp. 688–694, 2012.
- [34] M. Mohammed and N. Omar, “Question classification based on Bloom’s Taxonomy using enhanced TF-IDF,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1679–1685, 2018.