



Design and Implementation of an Item Bank Framework: A Developmental Study in National Medical Science Assessments

Abtin Heidarzadeh ¹, Parvane Parvasideh ², Zahra Bahrevar ², Hanieh Zehtab Hashemi ^{2,3}, ^{4,*}

¹ Medical Education Research Center, Guilan University of Medical Sciences, Guilan, Iran

² National Center of Medical Education Assessment, Ministry of Health and Medical Education, Tehran, Iran

³ Department of Health Informatics, Smart University of Medical Sciences (SmUMS), Tehran, Iran

⁴ Artificial Intelligence in Medical Sciences Research Center, Smart University of Medical Sciences, Tehran, Iran

*Corresponding Author: Department of Health Informatics, Smart University of Medical Sciences (SmUMS), Tehran, Iran. Email: zehtab@behdasht.gov.ir

Received: 28 October, 2024; Revised: 13 August, 2025; Accepted: 14 September, 2025

Abstract

Background: The increasing demand for standardized national medical exams necessitates the adoption of systematic, computerized question designs. Existing item banks often fail to meet the evolving needs of medical sciences, particularly in Iran, due to limited resources and the need for rapid updates, highlighting the need for a tailored framework.

Objectives: The present study aimed to design and implement an effective framework for item bank integration into national medical assessments.

Methods: This developmental study utilized document analysis to identify key criteria for item bank creation. Based on these findings, a comprehensive framework was developed and integrated into the national medical examination system in collaboration with the National Medical Education Assessment Center (NMEAC). Iterative reviews and adjustments improved item quality, accuracy, and alignment with assessment standards.

Results: The process involved designing test blueprints, managing exams, and improving item quality. The framework registered 900 professors and processed over 8,000 questions, with 6,000 reviewed for exam booklets. The system included feedback mechanisms and 24/7 support for question designers, with training conducted at the Smart University of Medical Sciences of Iran.

Conclusions: Item banks improve the quality and reliability of medical exams by ensuring secure systems and expert reviews. The distinction between item banks and pools ensures standardized test development and accurate measurement. Challenges such as Persian language compatibility and scalability remain, but the framework demonstrates potential for adaptation in other fields. Its principles may be applied to other disciplines with appropriate contextual adjustments. This study received no external funding or support. Its applicability beyond medical education may be limited without contextual adaptation.

Keywords: Item Bank, Question Banking, Test Construction, High-stakes Assessment

1. Background

The National Medical Education Assessment Center (NMEAC) in Iran, part of the Ministry of Health and Medical Education (MoHME), oversees medical exams, from question design to results analysis. To enhance efficiency, NMEAC recognized the importance of integrating electronic item banking and information technology in test development. Despite challenges

such as low-quality software and concerns about item confidentiality, adopting IT was crucial for streamlining test processes. The NMEAC aimed to implement a system that adheres to both technical and non-technical standards, ensuring efficient and secure test management (1).

Question banking stores exam questions with relevant metadata such as content area, learning outcomes, difficulty, and source. Questions are

organized using a catalog system, updated after use, and arranged by objectives to create exams. Key benefits include efficient test construction, quality assurance, transparency, and performance monitoring. Item banks support uniform teaching and assessment standards, while metadata aids administration. Exams are created through manual selection or computerized randomization, which helps prevent plagiarism but raises concerns about comparability and reliability, addressed by aligning questions with consistent standards (1).

An item in testing includes a question, instructions, response processing, and feedback, defined by IMS QTI as the smallest assessment unit. With the rise of e-assessment, diverse item types require item banks to manage complexity. Research focuses on item design for specific domains, addressing evolving needs in complex assessments. The system's key stakeholders — item designers, authors, administrators, translators, project managers, and psychometricians — handle item creation, management, and quality control. Essential use cases include creating, editing, and managing items. Nonfunctional requirements such as reliability, security, and scalability guide the system architecture to meet stakeholders' needs effectively.

In the literature related to item banks, numerous studies have focused on creating high-quality items, item selection methods, test construction, and security issues. However, there is less research providing information on item bank design and how items are collected (2). We seek strategies for effective item design and item bank utilization. Research indicates that item banks enhance test security and ensure consistent difficulty distribution, improving item performance and test design through streamlined writing, reviewing, and automated tools that incorporate statistical information.

The term "Item Bank", coined in the mid-1960s, was used to describe a group of test items "organized, classified, and cataloged like books available in a library" (3). Subsequently, Bruce Choppin and others interested in item banks based on the item-response theory (IRT) have attempted to distinguish between an "item bank" (a set of questions calibrated with an IRT measurement system and equal to a common scale) and an "item pool" (a set of items grouped by content but not calibrated). This distinction has not been widely welcomed, and the terms "item bank" and "item pool" are often used interchangeably today (4).

The distinction between an item bank and an item pool is crucial in medical education assessments. An item bank allows for standardized test development, precise difficulty adjustments, and improved measurement accuracy. In contrast, an item pool may lack the consistency needed for high-stakes testing. Given the sensitivity of medical assessments, a structured item bank ensures fairness, reliability, and alignment with educational standards. According to AMEE Guide No. 71, all test systems require an item bank and an administrator responsible for maintaining and classifying items according to the test blueprint. The administrator ensures the item bank remains current by communicating with the scientific committee about required question types and quantities. Factors affecting item bank size include test repetition, item usage, reusability policies, and whether answer keys are provided after tests (5).

Using a blueprint for test construction ensures alignment with educational objectives and allows for varying item difficulty levels. By employing a predetermined blueprint, tests maintain validity and reliability while covering essential concepts fairly. This approach prevents the assessment of irrelevant knowledge and avoids bias towards specific topics. A study by Sclater et al. found item banks typically contain between 500 and 1,025 items, with their size increasing over time, primarily influenced by financial resources rather than the item bank's age. This strategy optimizes resource use in the assessment process (6).

Random extraction from item banks often fails in medicine and health sciences due to rapid developments, leading to outdated items. Additional quality control is necessary to ensure appropriate item combinations and maintain the relevance of test questions (7). Examples of popular commercial software that provides comprehensive test construction services and includes question banks are FastTEST Web (8), Perception, Random Test Generator PRO, and Test Creator; also, TAO (9) is an example of an open-source question bank. The IMS International Question and Test Interoperability (QTI) standard for including metadata items is not met for some available item banks. These item banks face the challenge of interoperability with other item banks or online test systems. Some systems used pre-testing to calibrate items (10, 11), while others used the specialized title and subject of each item to label items with appropriate metadata (2).

A study conducted in 2004 on item banks showed that most of the studied item banks were implemented

in various ways using SQL databases and XML technologies (10, 12). McAlpine et al. developed an outline of the elements that constituted an item banking system, and the workflow and initial diagrams provided an overview of the overall system and user interaction (13). Chituc et al. developed an item bank for computer-based assessments, highlighting its role in service-oriented architecture (SOA). Their project encompassed item design, test construction, and analysis of reports, using unified modeling languages (UML) to represent key components of the system (14).

Nuntiyagul et al. developed a tool for categorizing items in an item banking system using feature selection. This tool classifies items lacking predefined keywords based on item text, and user satisfaction evaluations demonstrate the system's acceptable accuracy (15). Nguyen et al. proposed using the particle swarm optimization (PSO) algorithm to create tests that achieve specific difficulty levels. This method generates tests from item banks with varying difficulties, aligning item difficulty with user assessments from prior tests, demonstrating superior performance across most criteria (16).

Janpla and Wanapiron studied the requirements for designing a question-banking and intelligent test system, identifying five key modules: Ser management, question management, examination management, evaluation management, and scoring management, forming the system framework. The intelligent question banking system integrates five modules within a cloud computing framework, designed by the researcher to enhance functionality. Machine learning is applied to examination management, question management, and evaluation management modules. This enables automatic test classification, selection of appropriate items for test-takers, and automated assessment with feedback provided to examinees based on their results (17).

Item banks are crucial for educational assessments, especially in medical sciences, but research on optimizing item design strategies is limited. Existing studies focus on technical aspects, like item categorization, validation, and the integration of advanced technologies, with insufficient exploration of adapting these strategies to evolving medical education needs. In Iran, implementing a question bank for medical assessments requires addressing security, scalability, question quality, and cost management, alongside local challenges such as limited financial resources and slow updates to reflect new medical

knowledge. A tailored framework for Iran, aligning with both global best practices and local needs, is essential for success.

2. Objectives

How can strategies be effectively applied in the design of items within the framework of an item bank?

3. Methods

3.1. Study Design

This study employed a developmental research design to create, implement, and refine a framework for integrating item banks into national medical science examinations. A cyclical process of analysis, design, development, and iterative evaluation ensured alignment with the specific needs of medical assessments.

3.2. Document Study

Diverse sources — including textbooks, peer-reviewed articles, academic theses, NMEAC internal reports, and international standards — were analyzed. Sources were selected based on relevance, credibility, and alignment with study objectives, and the extracted information guided framework design and development.

3.3. Framework Development

Requirements were analyzed, and essential features were identified for integrating the question bank with the examination system. Automated and AI-based solutions were explored, and multiple modules were implemented to facilitate national exams. Evaluation considered question accuracy, diversity, difficulty, and participant feedback. Multi-stage expert reviews, iterative feedback, and structured validation procedures minimized bias and ensured consistency, accuracy, and relevance across all components.

4. Results

Key components of the test development cycle encompass building the item bank, designing tests, delivering tests, collating test data, analyzing test data, and importing statistics into the item bank (18). The Faradid web-based software was selected and tailored to meet the specific needs of the NMEAC after a thorough review of existing software. This software is based on a multi-topic shared model (19). Implementing Sadaf et

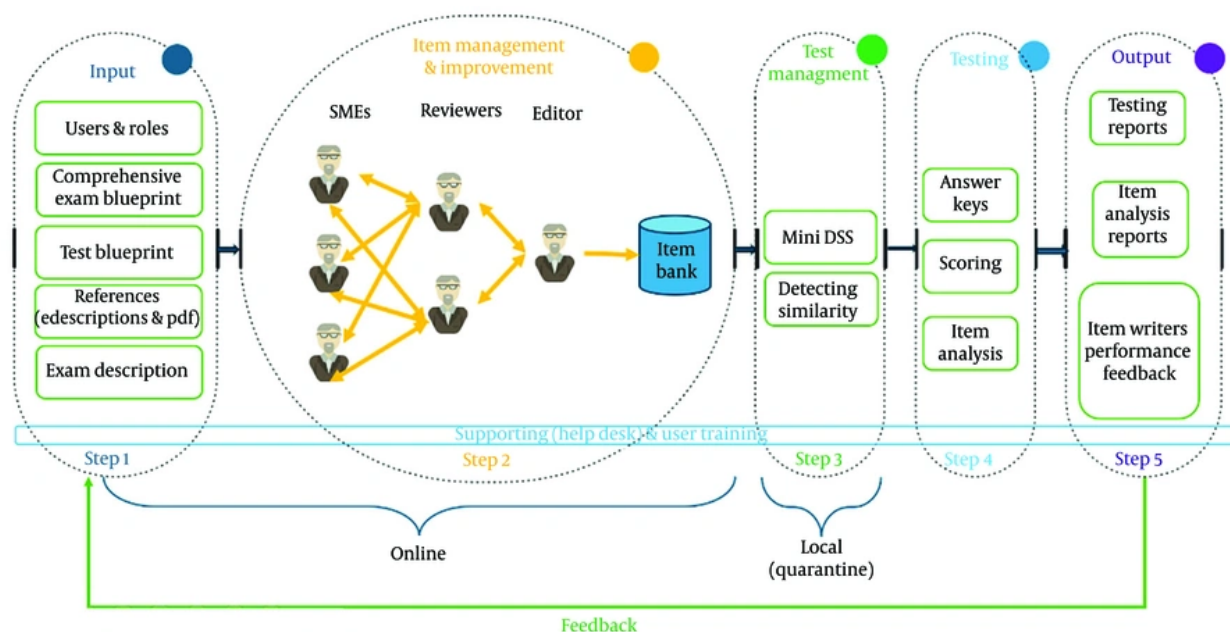


Figure 1. Architectures of an item bank system for national medical sciences examinations

al.'s step-by-step approach (20) ensured the development of a high-quality item bank that meets specific criteria for the assessment center. Modifications were made to adapt the process to medical science examinations.

This research utilizes components from Figure 1 to establish an item bank in both Web-based Item Banking Tools (WIT) and a quarantine environment. The web-based software supports various question types, including multiple-choice questions (MCQs) with 2 to 9 options, extended matching questions (EMQs), key feature (KF)/key feature problem (KFP) questions, true/false questions, and patient management problems (PMPs), with features for entering questions via an embedded word processor or importing from a Word file, along with plugins for complex formulas and diagrams. User log files are also recorded for retrieval.

4.1. Step 1: Input Users

- Various roles are defined, including administrators, typists, subject-matter experts (SMEs), reviewers, editors, and security users.

- Administrator responsibilities include defining users, access levels, and test-related resources.

- Typists handle corrections and modifications during the final stage of testing.

4.1.1. Preparing Comprehensive Blueprint and Test Blueprint

Comprehensive blueprints and test blueprints specify concepts for evaluation based on Bloom's taxonomy. Questions are gathered according to the comprehensive blueprint, and a table is utilized to allocate items for each test.

4.1.2. Uploading Resources' File and Users' Information

The system embeds the latest PDF files of required resources for professors, and users' information can be uploaded in groups using Excel format.

4.1.3. Exam Description

The exam management module allows adding, editing, deleting, and displaying exams, requiring the addition of majors, subjects, topics, and objectives beforehand.

4.2. Step 2: Item Management and Improvement

In the item management process, SMEs submit questions for review. First-level reviewers evaluate submissions and provide feedback if necessary. Approved items are sent to editors for a final quality check, ensuring alignment with educational goals and suitability for the target audience. Each question is assigned a score based on committee feedback, which influences its selection for the item bank, ensuring high-quality assessment items.

4.3. Step 3: Test Management

4.3.1. Selecting Questions for the Test

This step is performed in the quarantine environment with the presence of editors. Questions are selected based on the test blueprint and the determined proportions of the degree of difficulty. In the test booklet-making module, it is possible to randomize the items and questions. If the test is implemented electronically, on the test day, the selected set of questions is easily transferred to the test software in a safe environment. If it is used for a written test, it only needs to be printed and copied.

4.3.2. Detecting Similarities Between Questions and Mini Decision Support System

After final item design, software was created to detect question similarities in the booklet-making phase. Users can set similarity thresholds and apply filters. The system also checks for issues like option similarity and unusual question lengths, alerting users to approve or reject items that do not meet criteria.

4.4. Step 4: Testing Questions' Answer Key

The answer key construction for the questions is separate from and independent of the test. The test, after completion, is visible, can be printed, and can be downloaded as a file.

4.4.1. Item Analysis

This step involves reviewing questions and generating statistical information, such as difficulty and discrimination indices. Classical analysis and IRT are used to evaluate test psychometrics. If issues are identified, the problematic items are suspended until

reviewed. After each use, the indices are updated in the item bank.

4.5. Step 5: Output

- Testing reports and item analysis reports are generated, providing detailed insights into the performance of the test and individual questions.

- Item writers receive feedback on their performance based on the analysis reports, helping them understand areas for improvement and enhancing the quality of future question designs.

4.6. Challenges

Challenges during the project included ensuring question confidentiality. To address this, the process was divided into online and quarantine phases: The online phase allowed submissions, while the quarantine phase safeguarded confidentiality during selection. Additionally, many end-users were clinical specialists with limited computer skills, necessitating substantial helpdesk support.

4.7. General Specifications

- Security: All information is stored in an encrypted database, with no direct user access. Transactions occur through the system, which defines access levels based on user roles. Users can log in with a username, password, and one-time password (OTP) to access specific item bank sections and add new items.

- Usability: In designing the system, easy operation and graphic design were considered. Users could easily find out the number of required and designed questions for each lesson and each field on the page related to their desktop.

- Item writers training: A virtual modular training course consisting of 7 modules was prepared and implemented at the Smart University of Medical Sciences of Iran. This course aimed to familiarize participants with designing and assessing various types of questions in the field of medical sciences (21-23).

- Supporting: Virtual support of 24/7 ensured continuous communication and expert assistance, maintaining professor collaboration and effective implementation of the design process. The system registered 900 professors and received over 8,000 questions, with 6,000 reviewed for test booklets. Professors' departmental needs were addressed through meetings, leading to revisions like adding supervision

Table 1. Key Considerations for Integrating Item Banks into National Medical Science Examinations

Considerations	Details
Quality and validity	Item analysis, regular item rotation, maintaining psychometric properties
Development and maintenance	Investment in development, collaborative efforts to reduce costs
Technological integration	Computer-based testing, security measures
Educational impact	Focus on conceptual knowledge, use summative assessments
Practical considerations	Guidelines, review processes

levels in dentistry. An expert committee ensured quality assurance, promoting effective collaboration and maintaining high design standards for the medical test system.

This study presents a framework that not only enhances the quality of medical exam questions but also incorporates unique features such as high security, scalability, and clinical reasoning assessment. While previous studies (24, 25) have primarily focused on the technical design of item banks and summative testing models at the undergraduate level, this study emphasizes the specific needs of medical exams and offers continuous support for question designers through virtual training courses and 24/7 assistance. Compared to the I-MIB model for undergraduate evaluations (26), the proposed framework offers greater capabilities for more accurate and effective assessments. Considering these factors, item banks can significantly improve the quality, efficiency, and educational value of high-stakes national medical science exams. The considerations examined in this study are outlined in Table 1.

5. Discussion

This study examined the implementation of a computer-based item bank to enhance national medical examinations. Implementation involved question collection, multi-stage expert reviews, and final test assembly within secure online and quarantined environments. Collaboration among administrative leadership, technical support, and ongoing assistance for question designers was key to success. Technical challenges included Persian language processing, item version control, and data security, partly managed through parallel manual processes. Limitations involve limited generalizability to non-medical disciplines, scalability challenges for diverse content types, and the need to strengthen security with tools such as biometric authentication and digital signatures. Despite these challenges, integrating item banks provides clear

benefits, including precise knowledge assessment and flexible test design. While well-suited for medical sciences, application to other fields requires contextual adaptation. Future research should focus on improving scalability, enhancing security, and exploring broader applicability. These findings align with previous research on item bank development in high-stakes medical assessments, emphasizing security, scalability, and contextual adaptability (14, 15). Similar to the I-MIB model (26), this framework incorporates multi-stage expert reviews but differs in its explicit focus on clinical reasoning and bilingual interface support, highlighting that core item bank principles are transferable, yet domain-specific adaptations are essential for validity and reliability.

Footnotes

Authors' Contribution: Study concept and design: A. H. and H. Z. H.; Drafting of the manuscript: H. Z. H.; Critical revision of the manuscript for important intellectual content: P. P. and Z. B.; Administrative, technical, and material support: A. H.; Study supervision: H. Z. H.

Conflict of Interests Statement: The authors declare no conflict of interests.

Data Availability: No new data were created or analyzed in this study.

Funding/Support: This study received no financial support from grants or scholarships.

References

1. Sinharay S, Rijmen F, Choi S, Dorans NJ. The Revised Standards and Its Role in Research on Educational Measurement. *Educ Measur: Issues Practice*. 2014;33(4):36-8. <https://doi.org/10.1111/emip.12054>.
2. Baker EL. Introduction: Policy and technical Contexts of National Assessment of Educational Progress Validity Studies. *Educ Assessment*. 1995;3(1):1-8. https://doi.org/10.1207/s15326977ea0301_1.

3. Banerjee S, Rao NJ, Ramanathan C. Designing Item Banks in Alignment with Course Outcomes for Engineering Courses. *2016 IEEE Eighth International Conference on Technology for Education (T4E)*. 2016. p. 152-5.
4. Gierl MJ, Lai H. The Role of Item Models in Automatic Item Generation. *Int J Testing*. 2012;**12**(3):273-98. <https://doi.org/10.1080/15305058.2011.635830>.
5. Wrigley W, van der Vleuten CP, Freeman A, Muijtens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach*. 2012;**34**(9):683-97. [PubMed ID: 22905655]. <https://doi.org/10.3109/0142159X.2012.704437>.
6. Sclater N, Boyle E, Bull J, Church C, Craven P, Young R, et al. Defining the infrastructure for a national item bank service. *Ninth International Computer Assisted Assessment Conference*. 2025.
7. Schuwirth LW, van der Vleuten CP. How to Design a Useful Test. In: Swanwick T, editor. *Understanding Medical Education*. New York, USA: John Wiley & Sons; 2018. p. 275-89. <https://doi.org/10.1002/9781119373780.ch20>.
8. Sahin A, Hurtado Grooscors HA, Góngora-Cortés JJ. Review of FastTest: A Platform for Adaptive Testing. *Measur: Interdiscip Res Perspectives*. 2018;**16**(4):256-63. <https://doi.org/10.1080/15366367.2018.1492867>.
9. TAO Testing. *TAO Testing: Assessment Software*. 2025. Available from: <https://www.taotesting.com/>.
10. Zhu W, Fox C, Park Y, Fiset JL, Dyson B, Graber KC, et al. Development and Calibration of an Item Bank for PE Metrics Assessments: Standard 1. *Measur Physical Educ Exercise Sci*. 2011;**15**(2):119-37. <https://doi.org/10.1080/1091367x.2011.568366>.
11. Narayanan S, Kommuri VS, Subramanian SN, Bijlani K. Question bank calibration using unsupervised learning of assessment performance metrics. *International Conference on Advances in Computing, Communications and Informatics*. Chennai, India. 2017.
12. Zenisky AL, Sireci SG. Technological Innovations in Large-Scale Assessment. *Appl Measur Educ*. 2002;**15**(4):337-62. https://doi.org/10.1207/s15324818ame1504_02.
13. Liao W, Ho R. Framework and Implementation of the Virtual Item Bank System. *J Software*. 2012;**7**(6). <https://doi.org/10.4304/jsw.7.6.1329-1337>.
14. McAlpine M, van der Zanden L. Itembanking infrastructure: a proposal for a decoupled architecture. *Loughborough University*. 2006.
15. Nuntiyagul A, Naruedomkul K, Cercione N, Wongsawang D. Adaptable learning assistant for item bank management. *Comput Educ*. 2008;**50**(1):357-70. <https://doi.org/10.1016/j.compedu.2006.07.003>.
16. Nguyen T, Bui T, Fujita H, Hong T, Loc HD, Snasel V, et al. Multiple-objective optimization applied in extracting multiple-choice tests. *Eng Appl Artif Intell*. 2021;**105**. <https://doi.org/10.1016/j.engappai.2021.104439>.
17. Janpla S, Wanapiron P. System Framework for an Intelligent Question Bank and Examination System. *Int J Machine Learn Comput*. 2018;**8**:488-94. <https://doi.org/10.18178/ijmlc.2018.8.5.734>.
18. Weiss DJ. Item banking, test development, and test delivery. In: Geisinger KF, Bracken BA, Carlson JF, Hansen JJC, Kuncel NR, Reise SP, et al., editors. *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. Washington, USA: American Psychological Association; 2013. p. 185-200. <https://doi.org/10.1037/14047-010>.
19. Faradid Rayan Soft. [Faradid soft 2020]. 2025. FA. Available from: <https://faradidsoft.com/>.
20. Sadaf S, Khan S, Ali SK. Tips for developing a valid and reliable bank of multiple choice questions (MCQs). *Educ Health (Abingdon)*. 2012;**25**(3):195-7. [PubMed ID: 23823639]. <https://doi.org/10.4103/1357-6283.109786>.
21. Kohan N, Janatolmakan M, Rezaei M, Khatony A, Namaziandost E. Relationship between Learning Styles and Academic Performance among Virtual Nursing Students: A Cross-Sectional Study. *Educ Res Int*. 2021;**2021**:1-6. <https://doi.org/10.1155/2021/8543052>.
22. ShahAli S, Shahabi S, Kohan N, Ebrahimi Takamjani I, Ebrahimi R. Using e-learning methods for physiotherapy students learning – a systematic review and meta-analysis of the impact on knowledge, skills, satisfaction and attitudes. *Europ J Physiotherapy*. 2023;**25**(5):252-64. <https://doi.org/10.1080/21679169.2022.2085789>.
23. Smart University of Medical Sciences. [Sciences Suom, modular courses]. 2022. FA. Available from: <https://smums.ac.ir/%D8%AF%D9%88%D8%B1%D9%87-%D9%87%D8%A7%DB%8C-%D8%A2%D9%85%D9%88%D8%B2%D8%B4%DB%8C?id=35>.
24. Zhong S, Zhong Y, Li J, Wang W, Zhang C. The Design of Web-Based Intelligent Item Bank. In: Hui K, Pan Z, Chi-kit Chung R, Wang CC, Jin X, Göbel S, et al., editors. *Technologies for E-Learning and Digital Entertainment*. New York, USA: Springer; 2008. p. 278-89. https://doi.org/10.1007/978-3-540-69736-7_30.
25. Du QJ, Zou MG. The Design of Item Bank System of Computer Operating System. *Adv Materials Res*. 2013;**760-762**:2002-5. <https://doi.org/10.4028/www.scientific.net/AMR.760-762.2002>.
26. Janpla S, Piriyaawong P. The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates. *Univ J Educ Res*. 2020;**8**(9):4163-72. <https://doi.org/10.13189/ujer.2020.080942>.