[1] Puttachat Khuntontong

[2] Thanapapas Horsuwan

Towards Personalized Learning: Implementing Computerized Adaptive Testing for Tailored Educational Experiences

Journal of Electrical Systems

Abstract: - The integration of computerized adaptive testing (CAT) with the two-parameter logistic (2PL) model of item response theory (IRT) marks a significant advancement in personalized educational assessment, as showcased in "KlassBitsCAT." This innovative framework adapts in real-time to the examinee's ability, ensuring a testing experience that is not only tailored to the individual's performance but also grounded in rigorous statistical analysis and educational principles. The incorporation of question difficulty levels, inspired by Bloom's Taxonomy, adds a nuanced layer to the adaptation process, enabling the system to assess a wide range of cognitive skills from basic understanding to complex problem-solving.

Keywords— "Computerized Adaptive Testing", "Item Response Theory", "Two-Parameter Logistic Model", "Bloom's Taxonomy", "Educational Assessment", "personalized learning".

I. Introduction

The contemporary educational landscape is marked by a critical examination of traditional testing methods, which are often ill-equipped to cater to the diverse learning styles and paces of individual students. Traditional assessments, with their standardized and uniform approach, aim to evaluate student performance at a singular point in time, but in doing so, they may overlook individual progress and specific learning needs [1-5]. These conventional tests usually prioritize memorization and rote learning over critical thinking, problem-solving, and the application of knowledge in practical scenarios. Additionally, the pressure and stress associated with these assessments can adversely affect students' performance and overall well-being, as the fixed format of these exams does not support adaptive feedback or the personalized learning experiences essential for pinpointing and bridging learning gaps. There is an increasing acknowledgment within the educational community of the necessity for assessment methods that are more adaptable and reflective of each student's comprehensive abilities, learning journey, and potential.

In response to the shortcomings of traditional assessment strategies, adaptive testing methodologies, such as the KlassBitsCAT system, have emerged as innovative solutions designed to provide a more accurate and nuanced evaluation of student abilities. Unlike the conventional one-size-fits-all model, adaptive testing dynamically adjusts the complexity of questions based on an examinee's responses, thereby offering a tailored assessment experience. This approach not only enables educators to measure individual student performance more precisely but also enhances the potential for personalized learning and development. However, the transition towards such personalized assessments underscores the importance of having advanced technological frameworks in place, while also sparking discussions on the accessibility and fairness of deploying sophisticated educational technologies across diverse learning environments.

The "KlassBitsCAT" system takes a fresh approach to learning assessments by using Bloom's Taxonomy, a well-known educational framework that categorizes learning objectives into different levels of difficulty. This model cleverly maps out five levels of question difficulty, ensuring that each test is perfectly suited to a student's individual learning stage. It's like having a personalized guide for each student, helping educators pinpoint exactly where a student shines and where they might need a bit more help. By doing this, "KlassBitsCAT" doesn't just test what students know; it actively helps shape a learning experience that grows with them. This method is all about making learning personal, using technology to adapt to each student's needs. It's an exciting step towards making education more about understanding each student's journey, making sure that no one gets left behind just because they learn differently.

[1] Department of Robotics and Artificial Intelligence Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand. puttachat.kh@kmitl.ac.th

[2] Datability Co., Ltd., Bangkok, Thailand

II. Literature Review

The exploration of Computerized Adaptive Testing (CAT) models within educational contexts reveals a broad spectrum of applications aimed at enhancing the assessment process[6-11]. These models, which include the likes of the Two-Parameter Logistic (2PL) and Three-Parameter Logistic (3PL) models, are designed to adjust the difficulty of test items in real-time based on the test-taker's ability, thereby providing a more individualized assessment experience. In educational settings, CAT models have been applied to various disciplines, ranging from mathematics to language arts, demonstrating their versatility and potential to improve the accuracy and efficiency of student evaluations. Despite their widespread application, these models are continuously being refined to better address the unique challenges of educational assessment, such as reducing test anxiety, minimizing the effects of guessing, and providing immediate feedback to support learning.

Personalized learning strategies, underpinned by the principles of adaptability and learner-centered instruction, have shown considerable promise in enhancing educational outcomes. By tailoring the learning experience to the needs, preferences, and abilities of individual learners, these strategies aim to foster a more engaging and effective educational environment. Research in this area highlights the positive impact of personalized learning on student engagement, motivation, and achievement, suggesting that such approaches can significantly improve the learning experience. Furthermore, the integration of technology, including adaptive learning systems and data analytics, has further enabled the customization of instruction, allowing educators to pinpoint and address learning gaps more effectively. Despite these advancements, ongoing research continues to explore the optimal balance between personalized instruction and curriculum standards to maximize learning outcomes for all students.

Despite the advancements in CAT models and personalized learning strategies, gaps remain in current approaches that the "KlassBitsCAT" system seeks to address. One such gap is the need for CAT systems that are not only adaptive and efficient but also accessible and easy to implement within diverse educational settings. Additionally, while existing models have made strides in customizing assessments, there is still room for improvement in integrating comprehensive feedback mechanisms that can directly inform teaching and learning practices. "KlassBitsCAT" aims to bridge these gaps by offering a user-friendly, technologically advanced system that not only adapts to the learner's ability level but also provides actionable insights for educators to tailor instruction and support individual student growth. This focus on both assessment and instructional support distinguishes "KlassBitsCAT" from existing models and represents a significant step forward in the pursuit of truly personalized education.

Bloom's Taxonomy, a framework developed by Benjamin S. Bloom, categorizes educational objectives into a hierarchy that reflects the increasing cognitive complexity from simple recall to complex analysis and creation. This taxonomy serves as a foundational tool in education for designing curriculum, assessments, and instructional strategies, ensuring a holistic development of cognitive skills in learners [12-15]. It encompasses six cognitive domains: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, each representing a different level of cognitive processes. In the context of mathematics, as interpreted by Lindsey Shorser, Bloom’s Taxonomy aids in crafting questions and assignments that range from basic retention of terminology and facts (Knowledge) to the application of concepts in new situations (Application), and further to the creation of new mathematical ideas or theorems (Synthesis). This structured approach not only facilitates targeted learning outcomes but also supports educators in identifying and addressing the varied learning needs of students, promoting a deeper understanding and application of mathematical concepts. Through its comprehensive coverage of cognitive skills, Bloom's Taxonomy remains a pivotal guide in the design of effective educational experiences, aiming to foster critical thinking, problem-solving abilities, and creativity among learners.

III. Methodology

In the KlassBitsCat Model, a nuanced approach is employed to tailor the learning experience by evaluating the proficiency of learners across different dimensions, such as courses, grades, chapters, and specific learning objectives. The model initiates this process by selecting items at the intermediate difficulty levels—specifically, levels 2 or 3—out of a total of 5 difficulty levels. This initial selection is strategic, aiming to accurately gauge the learner's current understanding without overwhelming them with complexity or underestimating their knowledge. By starting at these levels, the model ensures a balanced assessment that can effectively challenge the learners while still being within their reach, thus facilitating a more precise measurement of their capabilities.

Following the initial item selection, the KlassBitsCat Model utilizes the 2PL (Two-Parameter Logistic) model to refine its assessment of the learner's abilities. The convergence of the 2PL model allows for a detailed and nuanced understanding of the learner's average level of proficiency concerning a particular learning objective. This adaptive methodology not only provides a dynamic assessment framework but also enables the creation of a personalized learning path that is closely aligned with the learner's unique needs and abilities. By systematically adjusting the difficulty of the learning materials based on the learner's performance, the KlassBitsCat Model fosters an environment of optimal learning growth, ensuring that each learner can progress at a pace that is both challenging and achievable.

Incorporating the Item Response Theory (IRT) models [16-19], specifically the Two-Parameter Logistic (2PL) and Three-Parameter Logistic (3PL) models, into Computerized Adaptive Testing (CAT) systems like KlassBitsCAT enhances the precision and adaptability of assessments. The 2PL model is expressed through the equation:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7a_i(\theta - b_i)}}$$

where P(θ) is the probability of a correct response, θ represents the examinee's ability, a is the item discrimination parameter, and b is the item difficulty parameter. This model recognizes that the likelihood of a correct answer is influenced by both the item's characteristics and the test taker's ability, allowing for a nuanced differentiation between varying levels of ability.

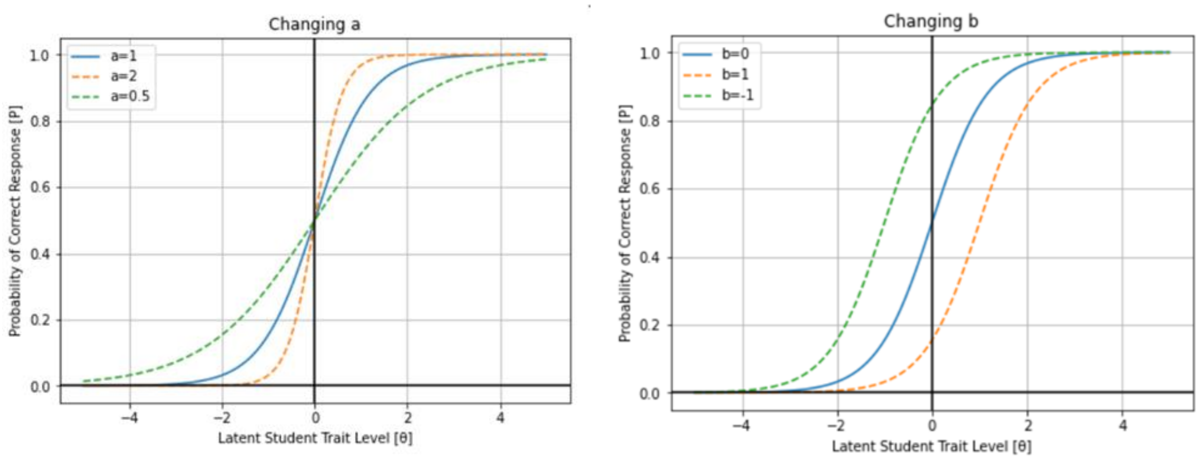
Conversely, the 3PL model adds a guessing parameter c, refining the equation to:

$$P_i(\theta) = c + (1 - c)\frac{1}{1 + e^{-1.7a_i(\theta - b_i)}}$$

Where θ is the ability level, a is the discrimination parameter, b is the difficulty parameter, and c is the guessing parameter, which accounts for the chance of guessing an answer correctly. This adjustment acknowledges the reality of test-taking strategies and the possibility of correct responses without full knowledge, thus providing a more comprehensive evaluation framework. However, despite its thoroughness, the 3PL model demands more extensive data for accurate parameter estimation and increases computational complexity. Opting for the 2PL model, KlassBitsCAT strikes a balance between detailed assessment and practical application constraints. This decision was made considering the operational efficiency and the model's capability to offer sufficiently precise insights into student abilities for most educational scenarios. Thus, the 2PL model was deemed an appropriate choice for KlassBitsCAT, emphasizing effectiveness and accessibility in educational assessments.

IV. Results

As IRT provides an estimation of the probability of a correct response to each question for a given value of examinee ability, a response with 2PL distribution is generated for each question by using the question parameters and the examinee ability to calculate the probability of correct response with changing parameters a and b.



(a) (b)

Figure 1. 2PL Model with Changing Parameters a and b

The exploration of the Two-Parameter Logistic (2PL) model in the "KlassBitsCAT" work provides insightful revelations into how varying the discrimination (parameter a) and difficulty (parameter b) impacts the model's performance and accuracy in assessing student knowledge. 5 difficulty levels from which simulated from Bloom's Taxonomy are Knowledge (-4), Comprehension (-2), Application (0), Analysis (2), and Synthesis (4). The simulations with changing a and b parameters underscore the nuanced relationship between item characteristics and test accuracy. When the discrimination parameter a increases, items are more capable of differentiating between students of different ability levels, leading to more precise estimations of student abilities. Conversely, adjustments to the difficulty parameter b allow for a more refined calibration of items to match the range of student abilities, ensuring that each test item contributes effectively to gauging a student's proficiency. This dynamic interplay between parameters highlights the model's flexibility and its capacity to adapt to diverse educational settings. The findings from these simulations not only validate the robustness of the 2PL model as a foundation for the "KlassBitsCAT" system but also emphasize its potential to enhance the precision of computer-adaptive testing in educational assessments. This adaptability is crucial for developing tests that are not only efficient and fair but also capable of providing meaningful feedback that can inform educational strategies and support individual learning pathways.

If the response pattern is locally independent, the test question of before does not effected the next question, the conditional probability of a single problem where an examinee with ability θ obtains a response on item I is denoted by $P(u_i \mid \theta)$. With u_i being a binomial variable of $u_i = 1$ for a correct response and $u_i = 0$ for an incorrect response, we get:

$$\begin{aligned} P(u_i \mid \theta) &= P_i(\theta)^{u_i}\left(1 - P_i(\theta)\right)^{1-u_i} \\ &= P_i(\theta)^{u_i}\left(Q_i(\theta)\right)^{1-u_i} \end{aligned}$$

Thus, the joint probability of an observed response pattern with n items can be denoted by the likelihood function: The estimate

$$\begin{aligned} P(u_1, u_2, \dots, u_n \mid \theta) &= P(u_1 \mid \theta)P(u_2 \mid \theta) \dots P(u_n \mid \theta) \\ &= \prod_{i=1}^{n} P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i} \end{aligned}$$

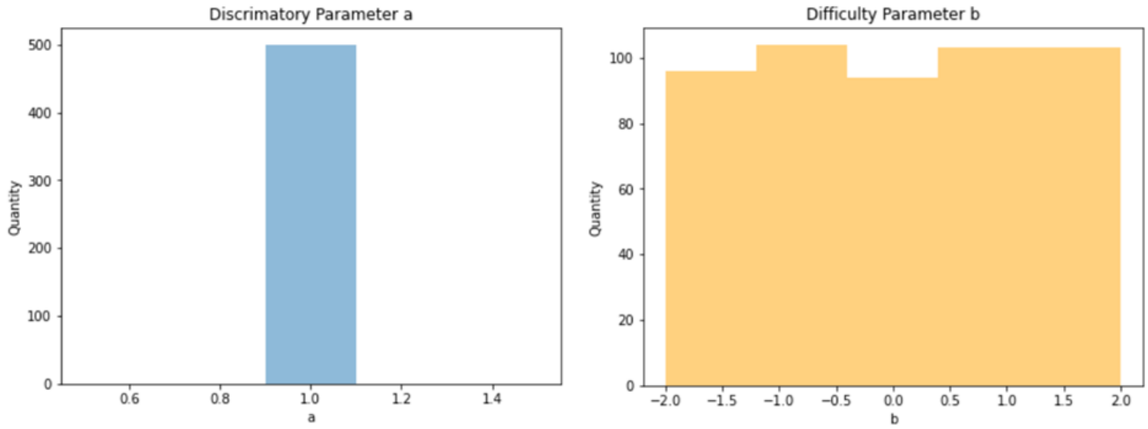
The log-likelihood function yields:

$$\begin{aligned} l(\theta) = \ln P(u_1, u_i, \dots, u_i \mid \theta) &= \ln \prod_{i=1}^{n} P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i} \\ &= \sum_{i=1}^{n} [u_i \ln P_i(\theta) + (1 - u_i) \ln Q_i(\theta)] \end{aligned}$$

Thus, the joint probability of an observed response pattern with n items can be denoted by the likelihood function. The Monte Carlo Simulation within the "KlassBitsCAT" system, particularly when examining the distribution of question parameters by varying a and b, reveals significant insights into the adaptive testing process. By systematically adjusting these parameters, the simulation demonstrates how the discrimination and difficulty levels of questions influence the accuracy and reliability of the assessment. For instance, increasing the discrimination parameter (a) tends to enhance the test's ability to differentiate between learners of varying proficiency levels more effectively. Meanwhile, adjusting the difficulty parameter (b) allows for the calibration of questions to better match the ability spectrum of the test takers, ensuring that each question is appropriately challenging. These findings are critical as they underscore the importance of precise parameter tuning in developing an adaptive testing system that can accurately measure student abilities across a wide range. The simulations not only validate the robust framework of the "KlassBitsCAT" model but also highlight its potential to optimize educational assessments through adaptive methodologies. This approach ensures that assessments are not just a measure of what students know at a fixed point in time but a dynamic process that evolves with the learner, offering insights that are pivotal for personalized education strategies.

In the "KlassBitsCAT" system, the Monte Carlo Simulation methodology, particularly when applied to the distribution of quality questions reveals significant insights into the adaptive testing process. By fixing the discrimination parameter a at 1 and allowing the difficulty parameter b to vary as a discrete random variable across five difficulty levels (-2, -1, 0, 1, 2, corresponding to Levels 1 through 5),

This finding indicates that by maintaining a constant level of discrimination while adjusting the difficulty levels, the test ensures that each question consistently contributes to accurately assessing a student's ability, regardless of its position on the difficulty spectrum. This uniformity in question quality is crucial for the integrity of the adaptive testing process, as it guarantees that every item has a fair chance of discriminating between different levels of student proficiency.



(a) (b)

Figure 2. Question Distribution. (a) Discrimatory parameter a, (b) Difficulty parameter b.

The simulation of the "StudentParametersDistribution" graph, Figure 3, the "KlassBitsCAT" model reveals a notable drop in student trial levels at levels 2 and 3. This finding suggests a concentration of difficulty or engagement issues that may impede learners' progression through these intermediate stages. To address this challenge, it is recommended to refine the question pool by enhancing the variety and adaptability of questions within these levels. Incorporating a broader range of difficulty and formats can help to better engage students, making the assessment more responsive to their varying abilities and learning styles. Additionally, offering targeted feedback and remedial resources at these critical levels could support learners in overcoming hurdles, ensuring a smoother transition and fostering a more effective learning trajectory.

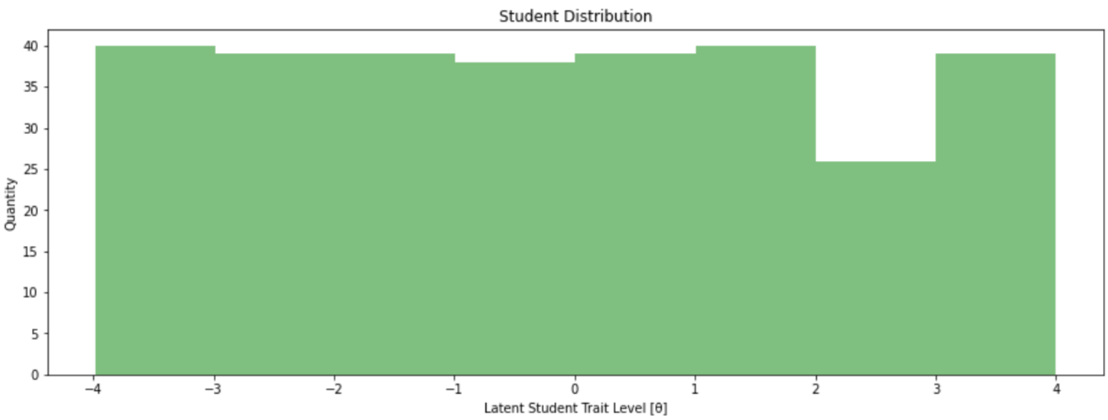


Figure 3. Student distribution along 5 level of questions difficulty and number of question required (quantity)

The color mesh graph produced by this code offers a compelling visualization of the dynamics between student abilities and question difficulties within a CAT system. It not only illustrates the practical application of the 2PL model in simulating adaptive testing scenarios but also provides an intuitive understanding of how well the test differentiates among varying levels of student ability. By sorting both students and questions before plotting, the graph highlights the effectiveness of the CAT in matching question difficulty to student ability, a crucial aspect in adaptive testing.

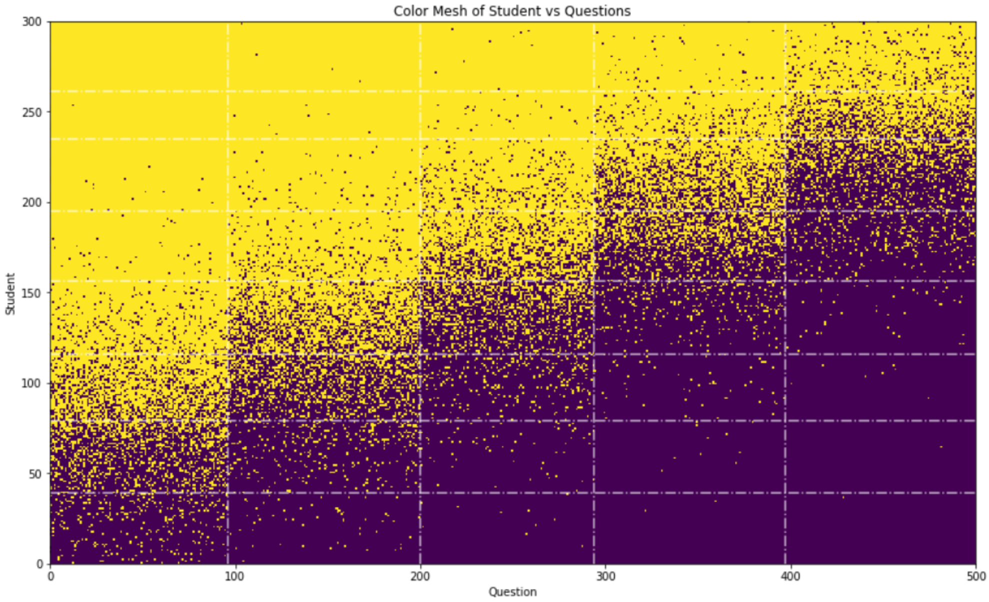


Figure 4. The color mesh which compelling visualization of the dynamics between student abilities and question difficulties.

V. Discussion

The "KlassBitsCAT" system incorporates Termination Criteria that play a crucial role in enhancing both the adaptability and efficiency of the testing process. This system employs a sophisticated approach to deciding when a test should be concluded, balancing the need for precise assessment of a learner's ability with the necessity of maintaining test efficiency. One pivotal parameter is the threshold for the standard error of measurement (SEM), which reflects the accuracy of ability estimates. A lower SEM signifies greater confidence in the learner's estimated ability level, with the system potentially adopting a predefined SEM threshold, such as 0.3 or 0.5 logits, to indicate sufficient precision in measurement.

To further refine this approach, the study suggests integrating a minimum items constraint, typically ranging from 15 to 20 questions per training or assessment, alongside a variable termination criteria that focuses on standard error termination with a value ≤ 0.315. This integration ensures that the test does not conclude prematurely, allowing for a comprehensive assessment of the learner's abilities while also adhering to the principle of testing efficiency.

Another critical parameter includes the maximum number of questions to prevent the test from becoming unduly lengthy and possibly leading to learner fatigue or disengagement. This limit guarantees that the test remains within a manageable duration, yet still aims for precise ability assessments. Moreover, the adaptation of the test based on the change in ability estimates—if consecutive questions do not significantly modify the learner's estimated ability, it indicates that a stable understanding of the learner's proficiency has been achieved.

Tailoring the Termination Criteria parameters to the specific educational context and objectives of "KlassBitsCAT" is essential. In scenarios of high-stakes testing, opting for a lower SEM to ensure higher precision may be preferable, whereas in formative assessments, a slightly higher SEM could be considered acceptable to shorten the testing time. These criteria are critical for finding the optimal balance between the accuracy of the assessment and its efficiency, rendering "KlassBitsCAT" a versatile tool suitable for a variety of educational environments.

VI. Summary

The "KlassBitsCAT" project's integration of Bloom's Taxonomy into a sophisticated CAT system represents a pioneering approach to educational assessment. By aligning test item difficulty with cognitive skill levels, the system offers a more nuanced and educationally meaningful measure of student ability and learning outcomes. This innovative framework not only adapts to individual student performance but also challenges them across a range of cognitive skills, from basic understanding to complex analysis and creation. Such a system not only enhances the precision and fairness of adaptive testing but also contributes significantly to our understanding of student learning processes and outcomes. Through detailed modeling, simulation, and visualization, "KlassBitsCAT" demonstrates how adaptive testing can be transformed into a tool for deeper learning and educational insight, making it a valuable contribution to the field of educational technology and assessment.

References

[1] S. Chen, B. Mulgrew, and P. M. Grant, A clustering technique for digital communications channel equalization using radial basis function networks, IEEE Trans. on Neural Networks, vol. 4, pp. 570-578, July 1993.

[2] Daniel O. Segall, Principles of Multidimensional Adaptive Testing, Computerized Adaptive Testing: Theory and Practice. Springer, Dordrecht, Ch3, pp. 57-70, 2010.

[3] Rotou, O., Patsula, L., Steffen, M., & Rizavi, S., Comparison of multistage tests with computerized adaptive and paper-and-pencil tests. ETS Research Report Series, Vol. 1, pp 1-27, 2007.

[4] James E. Carlson, Matthias von Davier, Item Response Theory, Advancing Human Assessment, Springer, pp. 133-178, 2017.

[5] Poh Hua, TAY, Joyce, LIU, Yick Chee, FONG, Using Computerised Adaptive Tests in Assessment for Learning, The 44th International Association for Educational Assessment (IAEA) Annual Conference, Oxford University, United Kingdom September 9 – 14, 2018

[6] Sari, H. İ., & Huggins-Manley, A. C., Examining content control in adaptive tests: Computerized adaptive testing vs. Computerized adaptive multistage testing, Educational Sciences Theory & Practice, Vol. 17, No. 5, pp. 1759–1781, 2017.

[7] Luo, X., & Kim, D., A top-down approach to designing the computerized adaptive multistage test: Top-down multistage. Journal of Educational Measurement, Vol. 55, No. 2, pp. 243–263, 2018.

[8] Ma, Y. C., Investigating hybrid test designs in passage-based adaptive tests, Doctoral dissertation, The University of Iowa, 2020.

[9] Martin, A. J., & Lazendic, G., Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience, Journal of Educational Psychology, Vol. 110, No. 1, pp. 27–45. 2018.

[10] Sari, H. İ., & Huggins-Manley, A. C., Examining content control in adaptive tests: Computerized adaptive testing vs. Computerized adaptive multistage testing. Educational Sciences Theory & Practice, Vol. 17(5), pp. 1759–1781, 2017.

[11] Thompson, N. A., & Weiss, D. A., A framework for the development of computerized adaptive tests. Practical Assessment, Research, and Evaluation, Vol. 16, No. 1, 2011.

[12] Bloom BS., Taxonomy of educational objectives: the classification of educational goals. New YorkNY, 1956.

[13] Anderson LW, Krathwohl DR., A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New YorkNY: Longmans , 2001.

[14] Peter Brusilovsky, Adaptive and Intelligent Technologies for Web-based Education, In C. Rollinger and C. Peylo (eds.), Special Issue on Intelligent Systems and Teleteaching, Künstliche Intelligenz, pp. 19-25, 1999.

[15] Anderson, L., & D. Krathwohl (Eds), A Taxonomy for Learning, Teaching and Assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition, New York: Longman, 2001.

[16] Brusilovsky, P., Methods and Techniques of Adaptive Hypermedia, in User Modeling and User Adapted Interaction, Vol. 6, 1996.

[17] Conejo R., Ed. Guzmán, E. Millán, M. Trella, J. L. Pérez-De-La-Cruz, A. Ríos, SIETTE: A Web–Based Tool for Adaptive Testing, International Journal of Artificial Intelligence in Education, Vol. 14, pp. 1-33, 2004.

[18] Weiss, D. J., Adaptive Testing by Computer, Journal of Consulting and Ginkal Pathology, Vol. 53, No. 6, pp. 774-789, 1985.

[19] Weiss, D. J., G. G. Kingsbury, Application of Computerized Adaptive Testing to Educational Problems, Journal of Educational Measurement, Vol. 21, pp. 361-375, 1984.

[20] M. A. Badr, A. I. Alolah and A. F. Almarshood,Transient performance of seriesconnected three-phase slip-ring induction motors, IEEE Trans. Energy Conversion, vol. 13, No. 4, pp 305-310, Dec. 1998