BOARD # 243: From Adaptive Testing to Adaptive Learning: An NSF IUSE project

Dr. Jason Morphew, Purdue University at West Lafayette (PPI)

Dr. Jason Morphew is currently an assistant professor at Purdue University in Engineering Education and serves as the director of undergraduate curriculum and advanced learning technologies for SCALE. Dr. Morphew is also affiliated with the Center for Advancing the Teaching and Learning of STEM and the INSPIRE research institute for Pre-College Engineering. Dr. Morphew's research focuses on the application of principles of learning derived from cognitive science and the learning sciences to the design of technology-enhanced learning environments. His research builds on design principles to examine the impact of educational technologies on student learning, interest, engagement, and metacognition in STEM.

Amirreza Mehrabi, Purdue Engineering Education

I am Amirreza Mehrabi, a Ph.D. student in Engineering Education at Purdue University, West Lafayette. Now I am working in computer adaptive testing (CAT) enhancement with AI and analyzing big data with machine learning (ML) under Prof. J. W. Morphew at the ENE department. My master's was in engineering education at UNESCO chair on Engineering Education at the University of Tehran. I pursue Human adaptation to technology and modeling human behavior(with machine learning and cognitive research). My background is in Industrial Engineering (B.Sc. at the Sharif University of Technology and "Gold medal" of Industrial Engineering Olympiad (Iran-2021- the highest-level prize in Iran)). Now I am working as a researcher in the Erasmus project, which is funded by European Unions (1M $_European Union & 7 Iranian Universities) which focus on TEL and students as well as professors' adoption of technology(modern Education technology). Moreover, I cooperated with Dr. Taheri to write the "R application in Engineering statistics" (an attachment of his new book "Engineering probability and statistics.")

Ben Van Dusen, Iowa State University of Science and Technology

From Adaptive Testing to Adaptive Learning: An NSF IUSE project

Abstract

Funded by the Improving Undergraduate STEM Education program of National Science Foundation, our project is focused on developing and implementing computerized adaptive testing (CAT) in a freely accessible online platform system named LASSO that encompasses several conceptual inventories across STEM. CAT is an adaptive assessment method that selection of test items based on students' real-time performance. This adaptive approach allows for precise and efficient measurement of student proficiency (sometimes also referred to as ability). By selecting questions at the appropriate difficulty level for students, the assessment system in LASSO can apply several algorithmic models to derive information about student skill mastery, content area learning, and student conceptual profiles. By developing an in-depth and detailed profile for each student, the adaptive testing system can provide instructors with individualized insights into student learning, which is particularly valuable for large enrollment introductory STEM courses where instructors are not able to collect this data in real time.

The core of our adaptive testing system uses Item Response Theory (IRT) and Cognitive Diagnostic Models (CDMs) to provide detailed analyses of student proficiency and skill mastery. Further, Transition Diagnostic Classification Models (TDCMs) offer the ability to develop conceptual profiles using the specific incorrect answers students select to identify misconceptions. These models offer a granular view of students' cognitive strengths and weaknesses and allows instructors to identify specific areas where students need improvement.

While adaptive testing provides instructors with a powerful tool for assessing students, large enrollment classes still present a challenge for providing in-the-moment instructional interventions. By integrating adaptive learning processes into an adaptive testing platform, our work aims to present a more complete framework for optimizing student outcomes in large enrollment STEM courses. This work-in-progress explores transitioning from CAT to adaptive learning. By leveraging the diagnostic insights from IRT and CDMs, we are developing an adaptive learning system that adaptively curates personalized learning pathways for each student. This system will select learning content and instructional materials tailored to individuals' skill mastery and ability. By integrating CAT with adaptive learning, we can create a continuous feedback loop where assessment informs instruction in real-time.

Introduction

The growing emphasis on personalized and adaptive learning has become a cornerstone of modern educational approaches. Adaptive systems tailor learning experiences to the individual strengths and weaknesses of students. At the same time, these systems strive to ensure fairness and avoid biases that may arise from societal norms embedded in instructional materials. Many adaptive learning systems use proscriptive algorithms for providing learning trajectories for students. For example, every student who does not master the questions in a lesson will receive the same remediation, whether it is an instructional video, a worked-example, or a new problem set. However, with recent advances in artificial intelligence (AI), these systems can be greatly improved by curating a large repository of research-based learning activities, then utilizing machine learning to select the appropriate learning material.

One of the primary challenges in these sort of AI-enhanced adaptive learning platforms is the effective selection and delivery of instructional content. The learning activities must align with both the specific learning objectives of the course and student traits, such as proficiency and skill mastery. However, this becomes a difficult problem to solve as each learning activity may cover different concepts and teach students using different skills. An additional problem occurs in large-enrollment courses where adaptive learning platforms may not provide instructors with the personalized feedback that instructors need to develop or locate appropriate learning interventions. These challenges make it impractical to apply the same psychometric frameworks used in currently available computerized adaptive testing (CAT) platforms to the adaptive selection of appropriate learning resources. Our research aims to address this gap by designing a practical model capable of managing the informational parameters necessary for selecting instructional materials after adaptive testing. The proposed model will ensure that selected materials meet students' non-mastery areas while also aligning with their ability levels. Furthermore, the proposed model is adaptable such that it can incorporate additional parameters, enabling engineering educators to tailor content selection to their diverse learning goals and evolving classroom needs. By addressing these requirements, this research seeks to bridge the current limitations in adaptive learning systems and advance the integration of equitable and personalized content delivery.

Our current project has developed and implemented a cognitive diagnostic computerized adaptive testing (CD-CAT) that is freely accessible via an online platform system [1],[2]. This online assessment platform encompasses several static conceptual inventories across STEM; however, the CD-CAT has been implemented for introductory physics courses [3]. CD-CAT is an adaptive assessment method that uses adaptive testing algorithms to select test items based on students' real-time performance [4]. This adaptive approach allows for precise and efficient measurement of student proficiency [5]-[8]. By selecting questions at the appropriate difficulty level for each student, the CD-CAT more accurately derives information about student skill mastery, content area learning, and student conceptual profiles. By developing an in-depth and detailed profile for each student, the adaptive testing system can provide instructors with individualized insights into student learning, which is particularly valuable for large enrollment introductory STEM courses where instructors are not able to collect this data in real time.

The core of the CD-CAT uses Item Response Theory (IRT) and Cognitive Diagnostic Models (CDMs) to provide detailed analyses of student proficiency and skill mastery [9],[10]. IRT models the relationship between item characteristics and student ability. IRT models use logistic regression to model the probability of a student answering correctly on each item on an exam using parameters that estimate the item's difficulty, discrimination, guessing, and slip. The outcome variable of these models is an estimate for an unobserved or latent construct typically referred to as student ability or proficiency, which is updated by each item a student answers. CDMs enhance this process by identifying the underlying skills or learning objectives that students have mastered. CDMs follow a similar framework to IRT yet incorporate an attribute classification approach [11],[12]. CDMs classify students' mastery levels based on their responses to each question that was tagged with the skills needed to solve each item. By finding the maximum probability of assigning each student to different classes, the mastery profile of the skills will be achieved. Transition Diagnostic Classification Models (TDCMs) offer the ability to develop conceptual profiles using the specific incorrect answers students select to identify student misconceptions [13]. TCDMs are an extension of CDMs that incorporates a transitional framework to track changes in students' misconception profile over successive assessments [14]. The combination of IRT, CDM, and TCDM offers a granular view of the cognitive strengths and weaknesses of students and allows instructors to identify the specific areas where their student need improvement.

One limitation of this approach is that the creation of CD-CAT platforms requires a large number of student responses (800 – 1000) across a variety of proficiency levels and a knowledge of the psychometric models [15]. Our project aims to address this limitation by creating an open and customizable platform that all instructors can use with their students. While adaptive testing provides instructors with a powerful tool for assessing students, large enrollment classes still present a challenge for providing in the moment instructional interventions at scale. By integrating adaptive learning processes into an adaptive testing platform, our work aims to present a more complete framework for optimizing student outcomes in large enrollment STEM courses. This work in progress explores the next step in our project, which involves transitioning from CD-CAT to adaptive learning. By leveraging the diagnostic insights from IRT and CDMs, we are developing an adaptive learning system that curates personalized learning pathways for each student. This system will select video-based content and instructional materials tailored to individual skill gaps according to their skill mastery profile and abilities. We aim for the outcome to be an engaging, time-efficient, and effective learning experience, with content tailored to each student's ability level and mastery profile. By integrating CAT with adaptive learning, we can create a continuous feedback loop where assessment informs instruction in real-time. This adaptability ensures that each student's learning path evolves according to their progress, leading to improved academic outcomes and a more personalized educational journey.

Methodology

This paper reports on the findings of two methods for selecting learning materials using student skill mastery information, as well as other features of the learning material such as expected difficulty or time required to complete. This becomes an optimization problem as each learning activity comes with trade-offs. For example, a quick activity may not meet the needs of advanced students or cover fewer learning objectives. We approach this problem as a multi-objective optimization task. In a simulation study we simulate the performance of 1,000 students on a 60-item assessment that is designed to evaluate 5 skills or learning objectives. The use of a simulation study allows us to simulate a set of students with known learning needs, so that we can evaluate the effectiveness of the models. We also present a simplified optimization model to provide proof of concept, while also allowing for additional parameters to be added in future iterations. We then apply two different optimization algorithms – using either Greedy Heuristic or Gradient Descent – to examine the effectiveness with which these algorithms can adaptively assign learning activities to the students based on the student proficiency and skill mastery. To evaluate the algorithms, we focused on ensuring comprehensive coverage of non-mastered skills while simultaneously minimizing the number of assigned tasks and total task time.

Table 1: Properties of Simulated Learning Activities

Learning Activity	Length (min)	Difficulty	Skills Covered	Learning Activity	Length (min)	Difficulty	Skills Covered
1	6.5	difficult	2	11	15.0	medium	2
2	12.6	difficult	5	12	15.0	medium	5
3	15.0	difficult	1	13	15.0	medium	4,5
4	15.0	difficult	1,2	14	15.0	medium	3,4
5	15.0	medium	3	15	6.6	easy	5
6	15.0	medium	4	16	15.0	easy	2
7	15.0	medium	5	17	15.0	easy	1
8	5.0	medium	5	18	5.0	easy	1
9	5.1	medium	3	19	15.0	easy	1,3
10	8.2	medium	1	20	7.8	easy	3,5

To create the set of instructional activities, we simulated a set of 20 learning activities that had an expected time-on-task between 5 and 15 minutes. In addition, the tasks were defined as easy, medium, or difficult to reflect the diverse student proficiencies, as well as the skill(s) the learning task covered. The details of the learning activities can be found in Table 1.

Results

In the student performance simulation, only 49 students demonstrated mastery on all five skills. Both algorithms correctly did not assign any learning activities to these students. For the remaining 951 simulated students, both the Greedy Heuristic and the Gradient Descent algorithms selected enough learning activities to address the skills that the students had not mastered. In other words, all students who had not mastered a skill were assigned a learning activity, which is positive. However, students are enrolled in multiple courses, so over assigning learning activities may impact their performance in other classes. When we examine the percentages of students who received only one activity per unmastered skill versus those who received more than one learning activity for a given unmastered skill (Table 2), we find the Greedy Heuristic algorithm appears to be the most efficient algorithm. In addition, for both algorithms there were some learning activities that were not selected due to their length, difficulty, or other factor. This result indicates that both algorithms are able to reject learning activities that are less optimal for the given parameters in the models.

Table 2: Coverage Results for Each Algorithm

	Greedy Heuristic	Gradient Descent
Appropriate Coverage	932	892
Over Coverage	19	59

Conclusion

This study is the first step in creating an AI-enhanced adaptive learning algorithm that can easily be tuned to be effective in multiple contexts. While we only examined skill coverage, difficulty, and time needed to complete the learning activity, the optimization algorithms can be easily modified to include other research-based factors that are relevant for selecting appropriate learning activities. We are currently in the process of extending this simulation study into an authentic context in a large-enrollment introductory STEM course where students complete CD-CAT assessments online and are assigned learning activities from an existing course repository to examine the impact on student learning and course performance.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Awards No. 2322015 and No. 2142317. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] J. W. Morphew, A. Mehrabi, B. VanDusen, J. Nissen, & H-H. Chang, Computer adaptive testing in LASSO platform for classroom assessment and self-assessment. In ASEE 2024 Conference Proceedings, Portland, OR, June 23-26, 2024. https://peer.asee.org/46534
[2] P. V. Le, B. Van Dusen, J. M. Nissen, X. Tang, Y. Zhang, H-H. Chang, & J. W. Morphew, Applying cognitive diagnostic models to mechanics concept inventories. Physical Review Physics Education Research, 21, 010103, 2025. https://doi.org/10.1103/PhysRevPhysEducRes.21.010103
[3] J.W. Morphew, J.P. Mestre, H-H. Chang, H-A. Kang, & G. Fabry, Using computer adaptive testing to assess physics proficiency and improve exam performance in an introductory physics course. Physical Review Physics Education Research, 14, 020110, 2018. https://doi.org/10.1103/PhysRevPhysEducRes.14.020110
[4] H. Wainer, *Computerized adaptive testing: A primer, 2nd ed.* Mahwah, NJ: Erlbaum, 1998.
[5] H.-H. Chang, Understanding computerized adaptive testing: From Robbins-Monro to Lord and beyond, in Handbook of Quantitative Methods for the Social Sciences, edited by D. Kaplan (Sage, Thousand Oaks, CA, 2004), p. 117–133
[6] D. J. Weiss, “Improving measurement quality and efficiency with adaptive testing,” *Applied Psychological Measurement,* 6, 473–492, 1982.
[7] A. Sahin and D. Ozbasi, “Effects of content balancing and item selection method on ability estimation in computerized adaptive tests,” *Eurasian Journal of Educational Research,* 69, 21-36, 2017.
[8] S.-Y. Chen, P.-W. Lei, and W.-H. Liao, “Controlling item exposure and test overlap on the fly in computerized adaptive testing,” *British Journal of Mathematical and Statistical Psychology,* 61, 471–492, 2008.
[9] H-H. Chang, Psychometrics behind computerized adaptive testing, Psychometrika 80, 1, 2015
[10] H. Bock, and R. D. Gibbons, *Item Response Theory.* Hoboken, NJ: Erlbaum, 2021.
[11] J. Liu, W. Tang, X. He, B. Yang, and S. Wang, “Research on DINA Model in Online Education,” in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, A. Mobasheri, Ed. Cham: Springer International Publishing, 2020, pp. 279–291.

[12] E. Thompson, A. Luxton-Reilly, J. L. Whalley, M. Hu, and P. Robbins, “P. Bloom’s Taxonomy for CS Assessment,” in *Proceedings of the tenth conference on Australasian computing education-Volume 78*, 2008, pp. 155-161.
[13] A. Mehrabi, and J. W. Morphew, “*Uncovering the cognitive roots of misconceptions in physics education for engineering students through transitional diagnostic models*”, In ASEE 2025 Conference Proceedings, Montreal, CN, June 23, in press.
[14] A. Schellman and M. J. Madison, “Estimating the reliability of skill transitions in longitudinal diagnostic classification models,” *Journal of Educational and Behavioral Statistics*, p. 10769986241256032, 2024.
[15] C. L. Hulin, R. I. Lissak, and F. Drasgow, “Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study,” *Applied Psychological Measurement*, 6, 249, 1982.