Educational Data Mining and Personalized Support in Online Introductory Physics Courses

Farook Al-Shamali
Athabasca University
farooka@athabascau.ca

Hongxin Yan
University of Eastern Finland
hongya@student.uef.fi

Sabine Graf
Athabasca University
sabineg@athabascau.ca

Fuhua Lin
Athabasca University
oscarl@athabascau.ca

ABSTRACT
Physics has always been a challenging subject for many students. Research also shows a gap between instructional goals and actual student learning in introductory physics courses. This study focuses on two online first-year courses that cover classical mechanics of the physics curriculum at an open university in Canada. Each of the two courses is developed around a textbook and includes a locally created study guide enriched with animated videos, dynamic diagrams, and interactive exercises. This study aims at introducing a simple feature to provide physics students with personalization based on their background knowledge and at examining students' interactions with the online course materials. Relevant educational data are compiled using checkpoint quizzes, self-reflection questionnaires, examinations, and log data collected through the learning management system (Moodle). In addition, peer faculty feedback is collected. Positive correlations are expected between regular learning behavior and engagement in personalized support and students' performance on examinations.

Keywords
Data mining, Learning analytics, Learning management system, Moodle, Introductory physics, Distance education, Online learning, Personalized support, Learning behaviours.

1. INTRODUCTION
Despite its significance as a foundation for modern technological achievements, physics is perceived as a challenging subject by many students. In 1987, a prominent physicist, Richard Feynman (1918 – 1988), suggested "that physics shouldn't be taught in high school because most of the teachers lacked a passion for the subject" [1]. Researchers at the time also pointed out an alarming gap between expected learning outcomes and actual student learning in introductory physics courses [2]. This old problem called for a reconsideration of the traditional approach to teach this important subject.

The argument surrounding physics education is especially relevant to the distance education (DE) model, which is witnessing a period of accelerated evolution, powered by advancements in digital technology. Despite challenges linked to the nature of DE, the flexible presentation format of online courses breaks some of the traditional barriers and allows for new possibilities. This leads to the question about effective instructional design features in introductory physics courses that cater to all students and provide successful online experiences [3-5].

Farook Al-Shamali, Hongxin Yan, Sabine Graf and Fuhua Lin "Educational Data Mining and Personalized Support in Online Introductory Physics Courses" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 561 - 564

An online course delivered through a learning management system (LMS) can provide a multitude of data and information related to students' interactions with the course materials. Knowledge obtained from mining and analyzing available data, in combination with plug-in adaptive learning systems, can be used to guide individual students to study more effectively and to improve the quality and presentation of the course materials [6-8]. For instance, recent studies suggest that one of the common (and apparently less productive) students' practices in studying physics involves the "cramming" of relatively large quantities of the subject matter during short periods preceding exams [5,9]. Also, Imhof et al. indicated a "negative relationship between prior knowledge test score and predicted learning progress" in physics modules [3]. Even though the investigated courses cater to adult learners, not all students can effectively acquire online learning abilities [10]. Such observations highlight how personalized and adaptive learning are potentially effective concepts in the design of online physics courses.

A major advantage of the traditional face-to-face (F2F) educational model is the direct student-teacher interaction, which permits the instructor to make timely adjustments to the subject matter and teaching style to ensure better students' engagement. The assumption here is that the instructor is sufficiently flexible to make the required accommodations, and the size of the classroom is reasonably small so that accommodating individual students becomes practical. In an online class, however, students interact less with a dedicated teacher but more with the LMS and the course materials. This is especially true in the asynchronous delivery model, where course content (typically) consists of rigid learning resources developed with the "one-size-fits-all" teaching concept. Such a delivery format does not take into consideration the "individual differences, personal needs and personal development" of all students [7].

Chaw and Tang found that students' use of the LMS is correlated with the service quality it provides [11]. Also, the quality of an online course should be enhanced when instructors are equipped with effective learning analytics and data mining tools [12,13]. In particular, proper utilization of educational data promises to facilitate effective personalized learning in online courses, including personalized feedback and recommendations for extra learning materials [8,14-18]. Such individualized support is particularly important in physics courses where conceptual understanding is typically constructed vertically using scaffoldings provided by essential mathematical tools. Therefore, physics students are expected to appreciate personalized learning environments that evaluate their progress, fill individual knowledge gaps, and sharpen specific math skills if needed [19,12].

Learning analytics (LA) and educational data mining (EDM) have been used for a range of applications, including personalized learning [20]. In this paper, we introduce a work-in-progress research project that uses LA and EDM to examine students' interactions with the course materials in two online physics courses. More specifically, the project introduces a simple and practical adaptive feedback module that can be easily integrated into the LMS. It provides a level of personalization based on students' background knowledge directed toward reducing the knowledge gap among a diverse student group.

2. PHYSICS COURSES

This study focuses on two first-year physics courses offered online (through Moodle) at Athabasca University in Canada. The first course is an algebra-based introductory physics and covers conventional topics in classical mechanics. It is considered among the top 50 high enrollment courses at the university, with effective annual registrations that exceed 400 students. The second course is physics for scientists and engineers, which is the calculus-based version of the first course. Both courses share a mandatory home lab component [21].

The textbook is an essential educational resource in a typical physics course. However, traditionally, the textbook is written with the conventional classroom in mind. Therefore, in DE, the study guide becomes an important component that guides the student through different learning activities and course assessments. In particular, the study guides for the two courses are designed to complement the textbook and provide additional reading (and audiovisual) materials related to each unit and lab experiment (see Figure 1). An important component of the study guide consists of detailed solutions to physics problems related to each unit in the course.

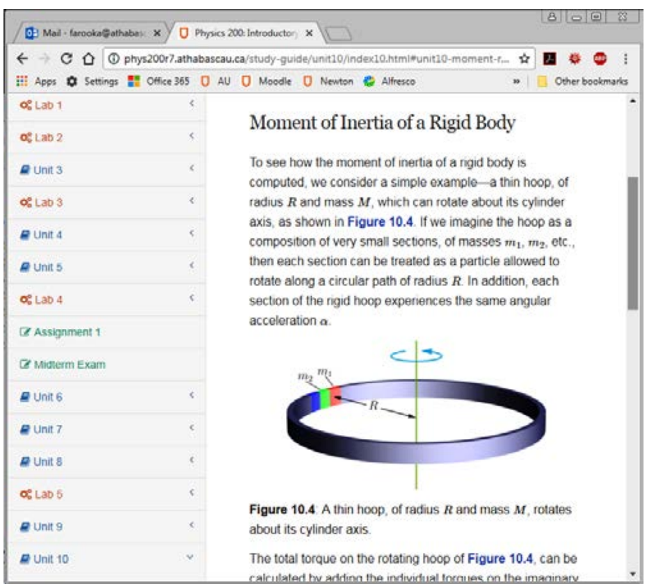

Figure 1. Snapshot from the introductory physics study guide.

Recent revisions of both investigated courses involved the production of an online study guide written in HTML with Mathjax scripting of the LaTeX code. The improved version of the study guide is enriched with dynamic and interactive diagrams created using Mathematica (https://www.wolfram.com/mathematica/). This is in addition to the free simulations of the PhET Project (https://phet.colorado.edu/). Students are expected to benefit from the interactivity and the real-time visualization of interactions between position, velocity and acceleration, especially in two and three dimensions. This includes the kinematics and dynamics of projectiles, circular motions, collisions, etc. Some of the interactive diagrams are complex enough to be considered virtual labs that simulate real-life situations.

The course development team constructed a website for the study guide that is accessible through the course homepage on Moodle and supporting responsive (mobile optimized) features. This is in addition to the textbook, which is accessible as an eTextbook through a separate link. One of the courses uses an open educational resource (OER textbook). The new design approach to the study guides received positive feedback from peer faculty members. However, even though some design considerations are integrated into the course for collecting feedback, our knowledge of students' interaction with the course content is limited.

3. RESEARCH Questions

In this study, we investigate the effectiveness of automated personalized support provided to students at specific milestones in two online physics courses. More specifically, the study addresses the following research questions:

- Is there a correlation between the academic performance of individual students and their response and behavior concerning the adaptive feedback module?
- How do learning behaviour and study patterns influence students' overall academic performance?
- What course elements are most effective regarding the adaptive feedback module?

4. RESEARCH PLAN

Relevant educational data are compiled using checkpoint quizzes, students' self-reflection questionnaire, course assessment results, and log data collected through the LMS (Moodle). In addition, peer faculty feedback will be collected.

4.1 Personalization through checkpoint quizzes

For the proposed personalization feature, the online study guide for each course is divided into five sections covering the main topics in each syllabus: kinematics, dynamics, energy & momentum, gravity & rotational motion, and elasticity & equilibrium.

Before starting a new section, a student is encouraged to complete a multiple-choice checkpoint quiz that is automatically marked by the LMS. The optional quiz is used as a checkpoint to assess the student's mastery of the topics in each section (see Figure 2). Based on the responses, the system may suggest the student proceeds to the next unit in the course or recommend a set of additional learning resources that may help strengthen the student's specific background concepts required by the upcoming topics. For example, a student who underachieved on the quiz questions related to rotational motion could be directed to a relevant video (such as https://youtu.be/garegCgMxxg) from the Khan Academy (https://www.khanacademy.org/), the problem-solving examples created in the study guide, or a section of the textbook. Apparently, there is limited research on "the effectiveness of such actionable links on students' learning experience and success" as stated by Iraj et al. [22]. The authors also warned that most students appear to lack "feedback literacy" and may only respond to quality feedback. This research project aspires to provide an informative contribution in this regard by using Moodle Quiz module's overall feedback feature, which can provide different feedback for a different level of quiz performance.

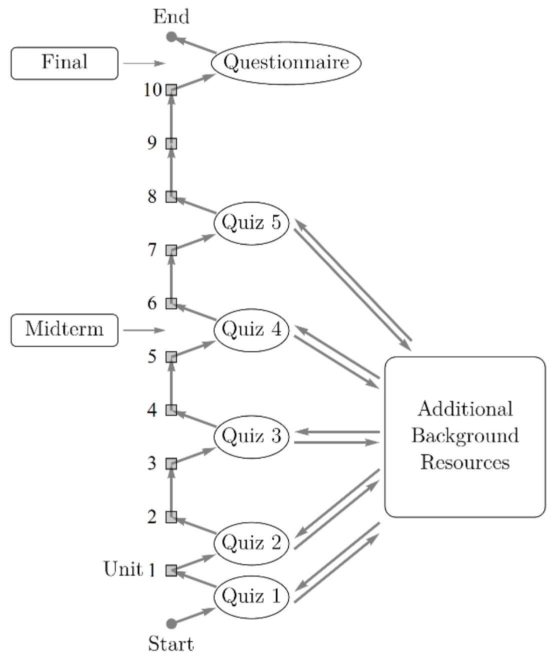

Figure 2. Learning path and checkpoints in the online physics course.

4.2 LMS Log Data

The LMS automatically records the date, time, and score of each checkpoint quiz and keeps track of other learning activities, including recommended learning objects accessed by individual students. When combined with the timing and marks achieved on assignments, lab reports and exams, each student's progress throughout the course materials can be highlighted. Such information allows us to look for patterns of effective learning behaviour and explore the effectiveness of the proposed adaptive feature on the student's academic achievement in the course.

4.3 Student Self-Reflection Questionnaire and Faculty Feedback

The effectiveness of the course content and design, in addition to the student's learning behaviour, is also gauged through a self-reflection questionnaire completed by the student toward the end of the course. The questionnaire is conduced online and consists of a mix of multiple-choice and written response questions. The collected data provides self-reflection by the students on their study behaviour, feedback on the proposed adaptive feature, convenience of course design, and effectiveness of course content, especially the interactive exercises and dynamic diagrams. Also, we will solicit qualitative assessment and feedback from instructors and tutors about the efficiency and effectiveness of the system through interviews.

4.4 Data Analysis

Based on the results of the first three checkpoint quizzes (see Figure 2), we group students by score quartile (students who perform below 25%; students with a score between 25% and 50%; students with a score between the 50% and 75%; and the students who score above 75%.) We then follow the learning behaviour of each group and their performance on the midterm examination. The fifth group of students who choose to skip the checkpoint quizzes, continuing from one unit to the next, can act as a reference group. A similar analysis is repeated for data collected during the second half of the course.

To compare the use of recommended learning resources across quartiles, we compute the mean number of resources accessed for each quartile. We hypothesize that students with higher exam scores tend to engage more seriously with feedback and follow a more regular study pattern (i.e., suggested study schedule) than those with lower exam scores. We will see if the response to feedback between student groups is significant at the $p<0.05$ level for different quizzes. The findings can be used to detect struggling students since they are less likely to use exercise for study purposes. Also, we conjecture that the students with the lowest grades have the lowest score on checkpoint quizzes and follow a more random study pattern. Their learning behaviour may be guessing or viewing hints in an attempt to build a catalog of correct answers, rather than actively using their knowledge to correctly address their knowledge gaps and adopt a more productive learning behaviour.

5. CONCLUSION

Students' interaction with the course materials, combined with their use of personalization through checkpoint quizzes, the results of self-reflection questionnaires, and peer faculty feedback, are analyzed in association with student's performance on assignments, lab reports and exams. The purpose is to look for educationally meaningful information regarding effective personalized feedback, successful learning behaviour, and good aspects of instructional design. An extension to this project would involve investigating the impact of personalizing the study schedule on the issue of procrastination and student attrition in online physics courses.

6. REFERENCES

[1] Hewitt, P. G., Ed. 2015. *Conceptual physics* (12th Ed.). Pearson Addison-Wesley, Petersburg, FL, USA.

[2] McDermott, L. C. 1990. Millikan Lecture 1990: What we teach and what is learned—Closing the gap. *American Journal of Physics*, 59, 4, 301–315.

[3] Imhof, C., Bergamin, P., Moser, I., and Holthaus, M. 2018. Implementation of an Adaptive Instructional Design for a Physics Module in a Learning Management System. In *Proceedings of the 15th International Conference on Cognition and Exploratory Learning in Digital Age* (CELDA 2018), 69–78. https://www.celda-conf.org/wp-content/uploads/2019/04/CELDA_2018.pdf

[4] Kortemeyer, G. 2016. Work Habits of Students in Traditional and Online Sections of an Introductory Physics Course: A Case Study. *Journal of Science Education and Technology*, 25, 697–703.

[5] Kortemeyer, G. 2019. It's All in the Data - But What is It? Learning Analytics and Data Mining of Multimedia Physics Courses. *International Journal of Physics & Chemistry Education*, 11, 1, 13–17.

[6] Dietz-Uhler, B. and Hurn, J. E. 2013. Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. *Journal of Interactive Online Learning*, 12, 1, 17–26.

[7] Peng, H., Ma, S., and Spector, J. M. 2019. Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment. *Smart Learning*

Environments, 6, 9. https://doi.org/10.1186/s40561-019-0089-y

[8] Khosravi, H., Sadiq, S., and Gasevic, D. 2020. Development and adoption of an adaptive learning system. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, (83-88). https://doi.org/10.1145/3375462.3375532

[9] Suhonen, S. J. and Tiili, J. A. 2015. Students' Online Activity on a Fully Online Introductory Physics Mechanics Course. In *Proceeding of the 43rd Annual SEFI Conference*, https://www.sefi.be/wp-content/uploads/2017/09/55889-SJ-SUHONEN.pdf

[10] Martinez, M. 2002. What Is Personalized Learning? The eLearning Developers' Journal, May 7, 1–8.

[11] Chaw, L. Y. and Tang, C. M. 2018. What Makes Learning Management Systems Effective for Learning? Journal of Educational Technology Systems, 47, 2, 152–169.

[12] CHEN, L. 2019. Enhancing Teaching with Effective Data Mining Protocols. Journal of Educational Technology Systems, 47, 4, 500–512.

[13] Martin, F., Ndoye, A., and Wilkins, P. 2016. Using Learning Analytics to Enhance Student Learning in Online Courses Based on Quality Matters Standards? Journal of Educational Technology Systems, 45, 2, 165–187.

[14] Armstrong, and Fuhua, L. 2010. Modeling and Personalizing Curriculum Using Petri Nets. In Proceedings of the 18th International Conference on Computers in Education, 10-12.

[15] Graf S., Kinshuk (2012) Personalized Learning. In: Seel N. M. (eds.) Encyclopedia of the Sciences of Learning. Springer, Boston, MA, 2592-2594.

[16] Greller, W. and Drachsler H. 2012. Translating Learning into Numbers: A Generic Framework for Learning Analytics. Educational Technology & Society, 15, 3, 42–57.

[17] El-Bishouty, M. M., Chang, T. -W., Lima, R., Thaha, M. B., Kinshuk, and Graf, S. 2015. Analyzing Learner Characteristics and Courses Based on Cognitive Abilities, Learning Styles, and Context. In Chang, M. and Li, Y. (eds.), Smart Learning Environments, Lecture Notes in Educational Technology, Springer-Verlag, Berlin Heidelberg, 3-25.

[18] Roberts, L.D., Howell, J.A. and Seaman, K. 2017. Give Me a Customizable Dashboard: Personalized Learning Analytics Dashboards in Higher Education. Tech Know Learn 22, 317–333. https://doi.org/10.1007/s10758-017-9316-1

[19] Bautista, G. 2012. The effects of personalized instruction on the academic achievement of students in physics. International Journal of Arts and Sciences, 5, 5, 573–583.

[20] Baker, R. 2016. Using learning analytics in personalized learning. *Handbook on personalized learning for states, districts, and schools*, 165-174.

[21] Al-Shamali, F., and Connors, M. 2010. Low-Cost Physics Home Laboratory. In Kennepohl, D. and Shaw, L. (eds.), Accessible Elements: Teaching Science Online and at a Distance, Published by AU Press, Athabasca University, 131-146.

[22] Iraj, H., Fudge, A., Faulkner, M., Pardo, A., and Kovanović, V. 2020. Understanding Students' Engagement with Personalised Feedback Messages. In Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20), March 23–27. https://doi.org/10.1145/3375462.3375527