LEARNING ANALYTICS TO MODEL LEARNING IN COMPUTER SCIENCE COURSE SOFTWARE

LEARNING ANALYTICS AND BARRIERS FOR INTERNATIONAL STUDENTS

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Carrington Murphy

October 31, 2019

Technical Project Team Members
Carrington Murphy

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: ______________________ Date: 11/1/19

Approved: ______________________ Date: Dec. 11, 2019
Catherine D. Baritaud, STS Division, Department of Engineering and Society

Approved: ______________________ Date: 12/10/19
Mark Floryan, Department of Computer Science

Over the past 30 years, readily available knowledge on the Internet has become one of the defining features of the educational landscape. Students have access to more learning content in today's classrooms than ever before. Researchers and policy-makers alike have sought to address the effects of this change on the learning process. As students are able to be more independent, self-regulated learning and virtual classrooms have become increasingly popular. One theory, personalized learning, is defined as an intentional and individualized approach to learning (Walker, 2019). This approach has been widely supported, due to the clear benefits of one-on-one teaching. It seems likely that increased individual time and support from a teacher will contribute to greater educational success. Unfortunately, in many scenarios this is not possible. Among the limiting factors are often large class sizes and access to the teacher. As the presence of technology in the life of students continues to grow, research continues to discover the ways in which computation and data can rectify some of the longstanding issues with education. Research in this paper will be centered on the use of learning analytics. For the purposes of this discussion, learning analytics refers to the leveraging of a given student's performance data to make predictions about their understanding of course material. The technical report will evaluate the degree to which learning analytics has a positive or negative impact on student success and retention. The STS paper, tightly coupled with the technical project, will use a social construction of technology framework to discuss barriers to student success and the ways in which learning analytics technology will increase or decrease the effects of those barriers.

LEARNING ANALYTICS TO MODEL LEARNING IN COMPUTER SCIENCE COURSE SOFTWARE

College students, specifically those enrolled in large lecture style classes, may often have a negative experience with their instructors. Bernard and Kuhn (2008) found a remarkably negative correlation between class size and student evaluations of instructor effectiveness. Many other studies have been conducted suggesting that students in large classes typically view their professors and classes as less instructive (Monks & Schmidt, 2010). This study hypothesizes that some negative results are at least partially due to a lack of individualized learning. Dale Basye, author of a book on personalized learning, defines individualized learning as an educational philosophy that seeks to tailor the pace of the curriculum to the individual student (Bayse, 2018).

In the Computer Science department at the University of Virginia, there is a core series of classes that is part of the curriculum for all students studying computer science. This study will examine one of these large courses, CS 2150, Program and Data Representation, which is a pre-requisite for most of the upper level electives. The material covered in this class builds many of the fundamentals for success in later courses. As a result, it is imperative that students are comfortable with the concepts taught in this course. In addition, personal communication with Professor Mark Floryan suggests that he believes many students are completing the course without clear understanding of all the course concepts (M. Floryan, personal communication, 3 September, 2019). In order to improve the course, Professor Floryan is building a piece of course software through which he hopes to encourage mastery of course material. In addition, the project aims to provide students with specific grading guidelines and live feedback. By using this system students may have better understanding of what is expected of them and their current standing in the class. This understanding may support greater student confidence and retention of material.

This technical research report will examine the ability of such a software system to accurately represent the learning process. In traditional class structures, an instructor will use student performance on quizzes, examinations, projects, and assignments to evaluate their knowledge of course material and assign a final grade. This project seeks to move this evaluation process into a computational model which predicts the probability that a student understands a given topic based on their performance in a series of opportunities to demonstrate that understanding. An effective system would allow students to progress through the class at their own speed while still encouraging mastery of all the course material.

Over the course of a two-semester capstone research project, under the advisory of Professor Mark Floryan, the individual technical project will analyze a variety of models which use learning analytics to provide essential feedback to both the student and the professor. The foundation of the application has been built by two project teams over the course of the 2018-2019 and 2019-2020 school years. This research project will implement its evaluation system on top of this existing application.

Specifically, this project hopes to expand upon a popular model originally developed by Corbett and Anderson (1995) called Bayesian Knowledge Tracing (BKT). The original model is a two-state model which is a function of four parameters, *P(L)* the probability that the student already knows some skill, *P(T)* the probability that the student learned the skill between the previous attempt this attempt, *P(G)* the probability the student does not know the skill but guessed it, and *P(S)* the probability the student knows the skill but slipped (Corbett & Anderson, 1995). The probability of the student knowing the skill *P(L)* is updated according to the two equations shown in Figure 1, depending on whether the skill was performed correctly or not. Once the probability reaches a certain threshold, the student is said to have demonstrated mastery of the skill.

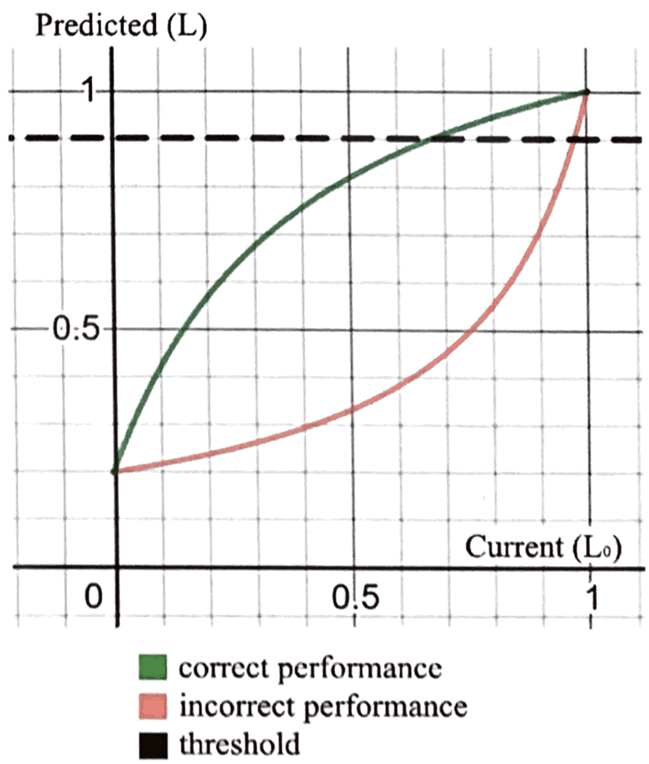

Figure 1: Conditional Probability in the BKT Model: This figure shows the two prediction functions developed in the classic BKT model proposed by Corbett and Anderson (1995). In addition, it shows the threshold for mastery-learning (adapted by Carrington Murphy from B. Van de Sande, 2013).

The classic implementation of BKT allows for the prediction of probabilities for a given topic. A classic interpretation of this system designed for an entire course would contain a series of these models which compute independently. This project will center on development and evaluations of extensions of this model which account for the way in which computer science classes rely on transfer of sequential knowledge. Analysis will center on a few key factors: statistical accuracy of predictions, student perceptions of instructor effectiveness, and overall success of the class in learning course material.

Research, development, and initial statistical testing will take place over the course of one semester. Afterwards, the system will be implemented as part of a course taught by Professor Floryan during which the course software will be tested further and analyzed for effectiveness. However, optimization of a large-scale software technology is an iterative process which will likely invite further exploration beyond the scope of this project.

Research and development will culminate in the production of a technical research paper discussing the process. This paper will be produced and submitted at the conclusion of implementation and preliminary study in the spring of 2020.

LEARNING ANALYTICS AND BARRIERS FOR INTERNATIONAL STUDENTS

The STS research paper will seek to further evaluate the effectiveness of learning analytics in encouraging student success. According to data from the Institute of International Education (2018), international students make up more than five percent of the total student enrollment in the United States. In addition, Araujo (2011) noted that, "findings from nine studies revealed that English fluency seems to be a significant variable related to the adjustment of international college and university students in the United States" (p.3). It is clear that learning and understanding English is one significant barrier for many students in higher education. In order to better understand these types of barriers, the paper will apply a framework exploring the social construction of higher education.

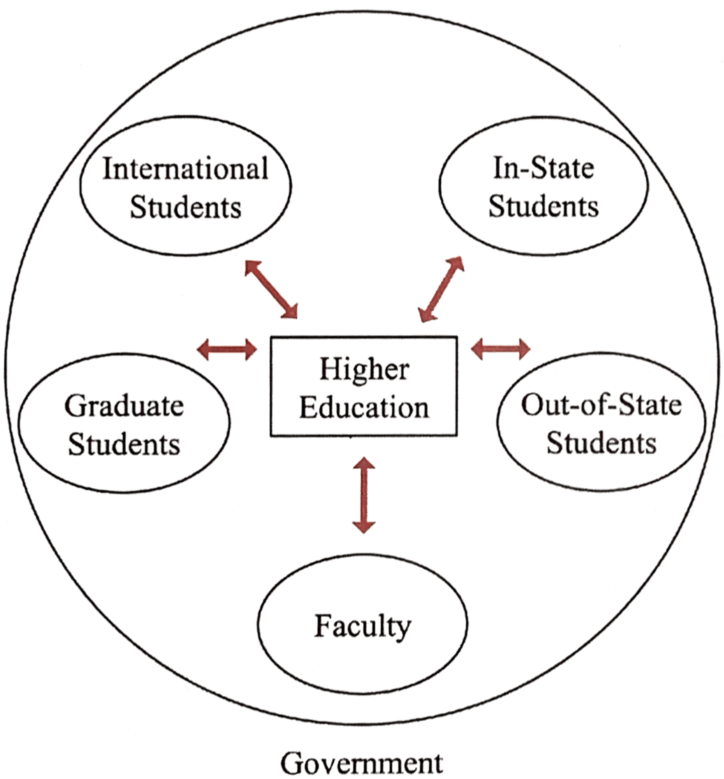

Figure 2: Social Construction of Higher Education: This figure shows some of the key social interactions between social groups and the technology of higher education. (adapted by Carrington Murphy from B. Carlson, 2009).

The existing context for this type of problem is further explored by Figure 2, which analyzes higher education as a technology. This social construction framework reveals two key insights. Firstly, higher education must play a distinct role in the experiences of different groups. For example, faculty need a higher education institution for their work while students expect a good education. The government also exerts pressure on the whole higher education process by determining regulations and funding for all groups involved. As such, it is clear that students, teachers, and the government all play an important role in the design of what higher education will look like. Secondly, an educational institution has a certain degree of flexibility in the way that it is thought of by certain groups. This is called interpretive flexibility. Students see often see a university as a short-term investment for learning and growth, while professors may rely on it for their life's work.

It should also be noted that the relationships between higher education and social groups can also be seen as an aggregation of the relationships between individual students and their education. The separation of these relationships is greatly supported by the educational philosophy of differentiated learning. Bayse (2018) describes differentiated learning as an educational philosophy that tailors the content of the curriculum and the methods of teaching to the individual student in order to maximize that student's success. Typically, differentiated learning can be associated with having greater quality and quantity of interactions between the student and the teacher. Growing interest in educational technology has created an opportunity for teachers to tailor their curriculum in new ways and provide many ways of learning material which were not previously available. Educational videos are an example of a technology that lowers barriers for all kinds of students because they are able to choose a time and place which will allow them to continue to learn the material even though they may struggle in the classroom (Kay, 2012).

IMPLEMENTATION AND ANALYSIS OF EDUCATIONAL SOFTWARE

The implementation of an educational software system should investigate the extent to which it is reinforcing or reducing language barriers for international students. This analysis will

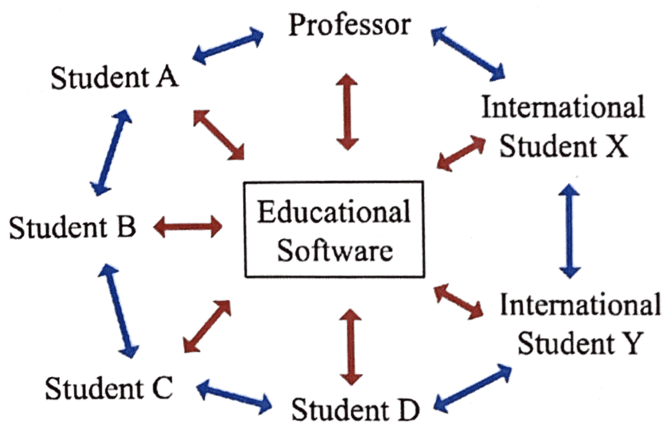

Figure 3: Technology and Social Relationships of an Educational Software System: This figure depicts the ways in which individuals interact with an educational software system with an additional focus on the relationships between individuals within the system. (adapted by Carrington Murphy from B. Carlson, 2009).

break down the social relationships in the classroom and the way that technology interacts with those relationships. The applied model, technology and social relationships, demonstrates the complex social network surrounding classrooms which should be analyzed further. Figure 3 depicts the key relationships in the classroom and with an educational software system. It is clear that some features of educational technology which are designed for average students, such as the language level used in instructions, may separately contribute to the success or failure of international students. In addition, it should be noted that the software system is only one part of a complex network of interactions among students and teachers. The arrows shown in blue are relationships that act independently of the technology. As such, the developers of a software should take care to recognize and attempt to strengthen those relationships, rather than simply replacing them.

The use of learning analytics technologies is somewhat controversial because of certain assumptions which are often made and may seem to place a lower value on the importance of the individual student. Perrotta and Williamson (2018), lecturers at the University of Leeds and the University of Stirling, respectively, critique the computational representation of a student as a "data double that can only ever be regarded as a temporary approximation stitched together from the available data points" (p.7). Perrotta's and Williamson's claims outline one of the key risks involved in using learning analytics: limited consideration of the many factors affecting learning. These concerns are especially valid for students who have limited understanding of English. For them, the learning process may not be able to be modeled in the same way. As an example, the classical model discussed in the technical report assumes that learning is a binary process and that it has a fixed probability of occurring each time a skill is performed (Corbett & Anderson, 1995). This discussion should bring about further investigation of the modelling techniques used by different algorithms.

Despite these criticisms, learning analytics technology may help to overcome significant language barriers by individualizing the learning process. By evaluating students on an individual basis, students have the ability to vary the pacing of their own education. The learning process is a complex one which takes different amounts of time for different people. The data representations created can actually allow for greater freedom and flexibility since students are not evaluated simultaneously. In a standard grading system, instructors usually make a short-term approximation of student understanding based on some subset of all the information about a student's background and experience. Examinations are typically the same for all students, and usually are taken at the same time. Learning analytics provides the ability to provide more specialization, where students are free to learn in the way that is best for them. While it is possible that learning analytics may not respond promisingly to all of the issues raised, it is important to recognize the potential benefits against the background of current methods. The classroom environment is complex, and learning analytics may only be part of a multifaceted solution to educational inequity.

The STS research paper will not only examine how learning analytics may affect the language barrier for international students, but how they perpetuate or destroy that which already exists. Doing so will require extensive research into the current issue and existing solutions, as well as the effectiveness of changes to the system. This research project will consist of an organized review of the ways in which learning analytics affects international learners. The report should seek to provide an outline of key questions and considerations to be made when implementing a course software.

WORKS CITED

Arujo, A.A. (2011). Adjustment issues of international students enrolled in American colleges and universities: A review of the literature. *Higher Education Studies*, 1(1), 1-8.

Bayse, D. (2018). *Personalized vs. differentiated vs. individualized learning*. Retrieved from https://www.iste.org/explore/Education-leadership/Personalized-vs.-differentiated-vs.-individualized-learning

Bedard, K. & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review*, 27(3), 253-65. https://doi.org/10.1016/j.econedurev.2006.08.007

Corbett, A.T. & Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model and User-Adapted Interaction*, 4(4), 253-278.

Institute of International Education. (2018). International student enrollment trends, 1948/49-2017/18. *Open Doors Report of International Educational Exchange.* Retrieved from http://www.iie.org/opendoors

Kay, R.H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820-831.

Molnar, A. (2019). *Virtual schools in the U.S. 2019 executive summary.* University of Colorado Boulder. Retrieved from https://nepc.colorado.edu/publication/virtual-schools-annual-2019

Monks, J. & Schmidt, R. (2010). *The impact of class size and number of students on outcomes in higher education*. Retrieved from http://digitalcommons.ilr.cornell.edu/workingpapers/114/

Murphy, C. (2019). *Conditional probability in the BKT model.* [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Murphy, C. (2019). *Social construction of higher education.* [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Murphy, C. (2019). *Technology and social relationships of an educational software system.* [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Perrotta, C. & Williamson, B. (2018). The social life of learning analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, 43(1), 3-16. doi: 10.1080/17439884.2016.1182927

Walker T. (2019). *Has the personalized learning hype worn off?* Retrieved from neaToday website: http://neatoday.org/2019/08/19/personalized-learning-research-raises-red-flags/

Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology.* Chicago, University of Chicago Press, 19-39.

Van de Sande, B. (2013). Properties of the Bayesian knowledge tracing model. *Journal of Educational Data Mining*, 5(2).