2019-06-04

Personalized Feedback: Testing a Tutoring System That Was Informed by Learning Analytics

McKenzie Emmett Staples
Brigham Young University, mckoo13@byu.edu

Personalized Feedback: Testing a Tutoring System That Was

Informed by Learning Analytics

McKenzie Emmett Staples

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Randall Spencer Davies, Chair
Gove Nathaniel Allen
Richard Edward West

Department of Instructional Psychology and Technology

Brigham Young University

ABSTRACT

Personalized Feedback: Testing a Tutoring System That Was
Informed by Learning Analytics

McKenzie Emmett Staples
Department of Instructional Psychology and Technology, BYU
Master of Science

Computer-based tutoring systems have been extensively studied and shown to be generally effective. In this study, I worked with an online beginner's course that teaches spreadsheets basics to business students. I designed four sets of practice problems that utilized a new tutoring feedback system that was designed for this course. In order to test the effectiveness of this feedback system, 839 Brigham Young University students that utilized this online course were randomly assigned to either a treatment condition where they worked through these new practice problems with the accompanying feedback system, or to a control condition, where they worked through the problems without the feedback system. Data about their performance on these problems, along with data about their performance on later assignments and a class midterm, were collected. ANOVA analyses showed that students in the treatment condition performed significantly better than students in the control condition in 17 of the 48 examined tasks ($p < .05$), with there being no significant difference in the other 31 tasks. These results indicate that the feedback system had a short-term benefit in many instances. Further research will be needed to determine if additional modification to the feedback system can extend these benefits to long-term situations.

Keywords: learning analytics, tutoring feedback system, online education

TABLE OF CONTENTS

LIST OF TABLES

DESCRIPTION OF REPORT STRUCTURE

This report is presented in a hybrid format containing two sections. The first section is in a journal article format that conforms to APA formatting guidelines. I anticipate submitting this manuscript to a journal that focuses on online education. The second section consists of an annotated bibliography that presents summaries of several articles pertaining to learning analytics and intelligent tutoring systems. It should be noted that some of the material written in the annotated bibliography is reused within the body of the journal-ready article. Per BYU's guidelines, this is permitted since the annotated bibliography will not be submitted for publication.

ARTICLE

Personalized Feedback: Testing a Tutoring System That Was Informed by Learning Analytics

McKenzie Emmett Staples

Brigham Young University

Introduction

Taking online courses is becoming a normal part of the college experience (Dennen, Darabi, & Smith, 2007; Picciano, 2012). Yet, just like traditional classroom courses, online courses need to be maintained and refined on a regular basis in order to improve their quality. One way in which online courses are being improved is through the use of learning analytics. Learning analytics in education has been described as the "collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2011, para. 3). Using data obtained from assessments and student interactions within an instructional system, researchers attempt to improve the ability of online courses to remediate student misunderstandings and provide actionable information to students and teachers so that student learning can be improved (Davies et al., 2017).

In this project, I utilized learning analytics data from the online course Prometheus Series: Microsoft Excel and Access (hereafter referred to as "this course"), which is hosted on the MyEducator platform. This course is used to teach students how to use Microsoft Excel and Access, though the portion I worked with covers only Excel.

In order to increase the likelihood that students would understand how to apply the material they learned in this course and to improve their performance on other assignments and tests, we (myself along with the others on the research team) augmented the structure of two lessons that already existed within this course. Prior to our alteration, each lesson in this course was broken into several segments. Each segment was accompanied by a set of practice problems that allowed students to practice the skills they had just learned. Accompanying each original set of practice problems, there was a complete step-by-step tutorial that students could use to see exactly how to complete each exercise. Additionally, at the end of each lesson, there was a final assignment (not accompanied by a tutorial) that tested the students' ability to combine and use all the skills learned in that lesson. Each end-of-lesson assignment was called a Test Your Skills (TYS) assignment.

Despite the fact that students are allowed up to two submissions on the TYS assignments, our initial data analysis of previous students revealed that many students would only submit each TYS once, even if they performed poorly. We believed that one reason for this was because students were expected to leap from a situation where they had complete help at every step of the assignment to a situation where they had to complete the problem entirely on their own without help. Thus, we believed that by adding a middle step with a tutoring feedback system, this would lead to an improvement in student performance since students would be able to do the work mostly on their own, but still receive some help when needed.

Purpose and Research Question

For this project, I designed new practice problems that were informed by a review of learning analytics data obtained from previous students who have taken this course. For students in the treatment group, these new practice problems were paired with a new tutoring feedback system that was designed by a fellow graduate student (Brice Colby) to enhance the feedback provided to students when they submitted solutions to practice problems (students in the control group received the new problems without the new feedback system). This allowed us to test whether students who had access to this new feedback system performed better than students who did not have access to it.

The specific question that we sought to answer with this research was: to what extent does the use of practice problems that provide diagnostic feedback based on learning analytics improve student learning?

Background Information

While teaching multiple students in a classroom is perhaps the most common type of educational setting, it has long been known that one-on-one tutoring is more effective (Bloom, 1984). One reason for this is because tutors are able to give timely and helpful feedback (Zhou, 1999). However, one of the chief reasons that mass education in a classroom is used rather than individual tutoring is that it requires fewer resources—it is easier to have one teacher instruct multiple pupils than to have one teacher for every pupil. However, with the advent of online schooling and other computer-based forms of instruction, this dynamic could change as students gain access to personal digital tutors. But do digital tutors allow students to experience benefits similar to the ones they experience with one-on-one tutoring? And is anything lost by removing the human interaction?

Kulik and Fletcher (2016) examined numerous studies that tried to determine the answers to such questions but that had conflicting results. They compared the results of these studies with previously available data about the effectiveness of human tutors and found that, in general, studies utilizing tutoring systems that are well implemented and robustly designed have effect sizes comparable to studies using human tutors (Kulik & Fletcher, 2016). Kulik and Fletcher also found that in many of the instances when tutoring systems were shown to be less effective, it was generally for one of two reasons, either (a) the tutoring system was poorly implemented (either because the instructors used the system as a replacement for classroom instruction rather than as a supplement to classroom instruction, or because the instructors were not trained in the use of tutoring systems) or (b) the control group used for comparison was poorly chosen (i.e., having other differences from the experimental group besides only whether or not the group received computer-based tutoring).

In order to construct computer-based tutors, designers typically utilize various types of student models (Chrysafiadi & Virvou, 2013). A student model is a system that uses available data about a student to extrapolate what the student may know or not know about the topic being taught by the computerized tutoring system; this model allows the tutoring system to adapt instruction and/or learning activities (such as practice problems) in order to meet the student's individual needs (Chrysafiadi & Virvou, 2013).

Most student models attempt to include what each student knows, how they may be using this knowledge, and what misconceptions about the material they have that may cause them to get wrong answers (Chrysafiadi & Virvou, 2013). However, the type of student model used in our newly designed feedback system is called a constraint-based model (CBM) and it is somewhat different than most other kinds of student models. According to Martin and Mitrovic (2006), CBMs work as follows: CBMs start by comparing a student's solution with a correct solution. The correct solution is split into its component parts (discrete pieces of knowledge or skills needed to solve the problem). Each component within the student's answer is checked to see if it is used correctly. If each of the knowledge components are used correctly, either nothing happens or the student receives a notification that she has provided the correct solution, depending on how the CBM feedback is set up. If any of the components within the solution are used incorrectly, the student receives a feedback message informing her of such until the mistake is corrected.

The format of feedback within a CBM can vary widely. Melis and Andrès (2005) outlined four different kinds of commonly used feedback: (a) “knowledge of result” tells a student whether her answer is correct or incorrect, (b) “answer until correct” tells the student when her answer is incorrect and prompts her to try again until it is correct, (c) “instruction-based elaboration” provides the student with information about the concepts she needs to understand in order to correct her errors and reach the correct solution, and (d) “knowledge of correct result” provides a student with the correct answer. The CBM used in our new feedback system utilizes the feedback types (b), (c), and (d) at differing times, depending on how much help the student requests from the feedback system (i.e., the system always uses type (b) by informing a student when her answer is incorrect and prompting her to try again; after two failed attempts, if she requests a hint, she is given feedback type (c); and if she requests a hint after three or more failed attempts, she receives feedback type (d)).

One further issue that should be kept in mind when implementing computer-based tutors is the tendency for students to try and game the system. According to Baker, Corbett, Koedinger, and Wagner (2004), gaming the system “consists of behavior aimed at obtaining correct answers and advancing within the tutoring curriculum by taking advantage of regularities within the software’s responses—systematically misusing the software’s feedback or help instead of actively thinking about the material” (p. 383). Research shows that students who game the system in courses tend to do more poorly overall compared to their peers (Baker et al., 2004).

However, Baker et al. (2004) also acknowledged that some techniques that may successfully prevent students from gaming the system may also make the tutoring system less effective for non-gaming students who need the extra help. Thus, designers should take care when considering whether to implement anti-gaming measures to make sure that they do not make the tutor difficult to use for the students who need it most. In order to address these two concerns, when our system was designed, it was designed so that it goes through a series of successively more helpful hints that apply to only one component of the specific task. Students must provide the correct response for that specific component before they are provided with any hints for another component within that task. Thus, any potential gamers would still have to do a large amount of work in order for the system to completely provide them with the correct answer (since that answer would be provided in piecemeal form), and students who need genuine help on a specific component within the task will still receive help they need.

Design and Methods

Preliminary Data Analysis and Assignment Design

Prior to the start of this study, I and another graduate student (Brice Colby) analyzed a data set containing the responses from over 4,000 students who completed the Lesson 2 and Lesson 3 TYS assignments in previous iterations of this course. First, we put these responses into two large spreadsheet files (one for each lesson). From there, we sorted them out to see which were the most frequently submitted answers. The amount of answers that we analyzed varied by each individual task, with anywhere from the top 18 most frequently used to the top 100 most frequently used answers being analyzed. The amount of answers analyzed depended on the number of students using each specific answer. Typically, the cutoff point was once five students or fewer submitted a specific answer, we stopped analyzing those answers, though in a few instances we raised this cutoff to as high as nine students per answer, simply for ease of analysis (in these instances of higher cutoff points, we still analyzed at least the top 100 answers per task).

Once we generated this list of top answers for every task, we then identified which of these answers were correct, incorrect, or partially correct. For each of the incorrect (and partially correct) answers, we identified why each response was incorrect. For example, most incorrect answers were either missing necessary components (such as using the incorrect formula), misusing necessary components (such as using incorrect syntax within the formula), or using unnecessary components (such as using absolute values when they were not needed).

After reviewing the material taught in these lessons, the current set of practice problems, and the common errors that we identified, I designed four new sets of practice problems: one each for two sections of Lesson 2 and one each for two sections of Lesson 3. The tasks within these new practice sets were designed so that they would be likely to elicit some of the common errors that were identified in the preliminary data analysis. We wanted the students to be likely to commit these errors so that they would trigger the newly designed tutoring feedback system which would help them work through the problem (and thus provide us data about their performance).

In order for the new problems to elicit errors from students, I examined the problems in the TYS assignments (i.e., the problems the students had worked with that provided data for our original data set), as well as the original set of practice problems for the sections in question (the ones that were accompanied with step-by-step guides). I then created new assignments that required students to perform similar tasks, but which were also situated in new contexts. For example, one task that students must complete in the Lesson 3 TYS assignment requires them to use the AVERAGEIF function to determine the average amount a customer spends per sale at several specific store locations. In the new practice problems that I created for Lesson 3, one task requires the students to use the AVERAGEIF function to calculate the average rating from critics that a film gets, based on which studio released it. Because these new problems were similar to the ones that originally elicited errors (but also dissimilar enough that students would still have to apply what they had already done to a new context), we anticipated that they would elicit similar errors, thus allowing them to trigger the feedback system to help them through the problem.

Design of the Tutoring Feedback System

After I designed these new problems, Brice Colby designed a tutoring feedback system that would be triggered whenever a student would commit an error. When an error is committed, students would be notified that they committed an error and would be offered a hint. Students would then choose whether or not to accept the hint.

The feedback system was designed so that each possible error could receive three levels of hints. The first time that a student commits an error for a specific component (for example, if they use the incorrect range) the first hint provided is very broad (identifying only where the error was committed), the second hint is provided when a student revises their work within that component, but still fails to get the correct answer. This second hint provides instruction on how to correct their mistake. The third and final hint is provided after the third failed attempt (as well as all subsequent failed attempts) to provide a correct answer within the context of a component. This hint would tell them exactly what the answer was for that component.

For example, if a student typed in =SUM(A1:A121), but the correct response was =SUM(A15:A121), the first hint offered would state that he used the incorrect range within his formula. The second hint would offer some instruction about how to determine the correct range to use (often pointing them back to the reading material in the lesson where they could learn more). The third and final hint would simply tell him to use the range of A15:A121.

Implementation and Testing

Once these new practice problems, along with the accompanying tutoring feedback system were fully designed, they were integrated into the online course.

For each of the four practice problem assignments, the students were assigned to either the control condition (receiving the new homework without the accompanying feedback system) or the treatment condition (receiving the new homework with the accompanying feedback system). The randomization process was designed so that each student would be in the control condition for two of the assignments, and the experimental condition for the other two assignments. Which group they were assigned to was determined by whether a specific digit of the student ID number was odd or even (i.e., all of the "odds" were part of the experimental group for two of the assignments, and part of the control group for the other two assignments).

Participants

The participants consisted of 839 BYU students enrolled in this course for the Fall 2017 semester. Students signed an IRB consent form from ongoing research with the MyEducator platform allowing for collection and analysis of their data. Students were informed that they could choose not to sign the form if they did not want their data used and that it would not affect their standing within the class if they chose not to sign.

Analysis

As students worked through the course, we collected student performance data from the four new homework assignments, as well as from the TYS assignments, and from the midterm test. When examining data from the TYS assignments and the midterm, we only looked at problems where students were required to use the various formulae that they learned in the sections we created new practice problems for. We ignored any problem in a TYS assignment or the midterm where students were required to use a formula learned outside of the four sections for which we created practice problems. Once this data was collected, we determined whether students were in the control or experimental group when a specific formula was taught, and we examined their performance on these problems to determine whether the condition they were in had an effect on their performance. For example, if a student was in the treatment condition when they learned the SUMIF function, we looked at their performance in every problem within the practice problems, TYS assignments, and midterm where they were required to use the SUMIF function to determine if being in the treatment condition had an effect. Overall, this selection process led us to extract performance data for 48 total problems across the practice problems, TYS assignments, and midterm examination.

Once the data for these 48 problems were extracted, we ran 48 one-way ANOVAs for each of the problems in order to compare three groups (which we labelled 0, 1, and 2.) Group 0 consisted of the individuals who were in the control group when a skill was taught (and thus did not have access to feedback system). Group 1 consisted of individuals in the treatment group who did not choose to view the hints they were offered, and Group 2 consisted of the individuals in the treatment group who did view hints when they were offered. It should be noted that this analysis only included instances when students committed an error. Thus, Group 0 consisted of students who were not offered any hints (due to being in the control condition) but would have been offered a hint had they been in the treatment condition. While the other two groups consisted of students who were offered hints and either viewed the hint (Group 2) or chose not to view the hint (Group 1). Thus, the overall N varied for each individual analysis, ranging from as low as 91 students to as high as 651 students. Prior to conducting the ANOVAs, we determined that p values of less than .05 would be considered significant.

Results

After analysis, we found that for nearly one third of each of the individual tasks examined, those who received help from the feedback system performed significantly better than both students who were not offered help and those who did not accept offered help from the system. However, we also found that most of the benefits from the feedback system occurred only within the new practice problems and that, with a few noted exceptions, there was not a significant difference on TYS assignment performance and midterm performance between the control and experimental groups.

Of the 48 problems we examined, we found that there were statistically significant differences between the three groups in 17 of the problems. Of these 17 problems, 14 came from the new practice problems, 2 came from the TYS assignments, and 1 came from the midterm exam. The data for each of these 17 statistically significant problems can be found in the Table 1.

Table 1

Problems with Statistically Significant Results.

Problem ID	Formula Tested	ANOVA *p* value	*p* Values for Comparing: Group 0 vs 1	Group 0 vs 2	Group 1 vs 2	Significant Relationships
2.2-C4	COUNTA	<.001	.035	<.001	.987	1>0, 2>0
2.2-C6	COUNT	.034	.818	.025	.429	2>0
2.2-C8	SUM	.017	.332	.015	.655	2>0
2.2-C10	SUM	.002	.119	.001	.686	2>0
2.2-C18	MIN	<.001	<.001	<.001	.130	1>0, 2>0
2.3-C7	FV	<.001	.002	<.001	.883	1>0, 2>0
2.3-C14	NPER	<.001	.454	<.001	.584	2>0
2.3-C15	EFFECT	<.001	.155	<.001	.001	2>0, 2>1
2.3-C22	PMT	<.001	.317	<.001	.247	2>0
2.3-C29	RATE	<.001	.898	<.001	.002	2>0, 2>1
2TYS-C23	RATE	<.001	<.001	.571	<.001	0>1, 2>1
2TYS-D37	PMT	.047	.240	.280	.046	2>1
3.2-E10	IF	<.001	.330	<.001	.003	2>0, 2>1
3.2-G10	IF/AND/IF	<.001	.959	<.001	.175	2>0
3.3-H7	COUNTIF	.003	.482	.002	.974	2>0
3.3-I7	SUMIF	.002	.167	.031	.004	2>0, 2>1
Mid-I16	AND	.023	.018	1.000	.019	0>1, 2>1

Note. Group 0 consisted of students in the control group who committed an error, and thus would have been offered a hint had they been in the treatment group. Group 1 consisted of students in the treatment group who did not choose to view the offered hints. Group 2 consisted of students in the treatment group who did choose to view the offered hints. The "Significant Relationships" column indicates which group(s) scored significantly higher than another group in that situation. For example, "2>0" indicates that for that problem in question, Group 2 score significantly higher than Group 0.

Discussion

These results indicate that our feedback system had the strongest effect in the short-term. That is, it helped students the most when they were first learning to develop their skills in the practice problems. This supports additional research that shows that feedback systems can be helpful when students are learning new material, as summarized by articles such as the literature review of tutoring systems done by Keuning, Jeuring, and Heeren (2016). They also identified in their article that feedback that helps students know how to proceed may be one of the most helpful types of feedback, which further supports our work since our feedback focused on helping the students know what to do next.

While we did see short-term results from our feedback system, there were only a few instances where significant long-term results were observed on items from the TYS assignments or midterm exam. One potential reason for this is because the alterations we made (with the new practice problems and the new feedback system) were implemented in an early portion of the course, which tended to have a lower overall difficulty level than the later parts of the course. Additionally, students may have learned the material from Chapters 2 and 3 in other ways as the course progressed. For example, when they began to see that they would need to continue to use the material that they were supposed to have learned in those chapters, they may have gone back and studied those chapters in order to deepen their understanding of the material. Likewise, as students progressed through the course and learned more difficult material which required them to employ the material that they had already learned, this additional practice may have naturally helped them to understand the material better. Since these possibilities are something that individuals in both the control and treatment groups would be exposed to, such an effect could level the playing field and potentially account for the general lack of long-term effects that we observed.

Future Research

It is possible that long-term effects could have occurred if we had focused our efforts on the later (and more difficult) parts of the course. This is something that could potentially be the subject of future research. Additionally, there are a couple interesting things to note about the situations where the feedback system *did* seem to have a long-term effect. In two of those three long-term situations (item 2TYS-C23 and item Mid-I16), it appears that the individuals who were in Group 1 (hint offered but not viewed) when those skills were taught scored not only significantly lower than those who viewed the hints, but also significantly lower than those *who were not offered any hints*. In these situations, it could suggest that those who chose not to learn from the offered hints missed out on key learning opportunities that may have caused them to perform more poorly on a later test, or it may merely indicate that those who choose not to view the hints even if offered tend to be the individuals who get lower scores. To determine which direction this relationship exists in (not viewing hints causes lower scores, or lower-scoring individuals choose not to view hints) will also require further research. One possible direction that this research could take may include some sort of qualitative inquiry that seeks to understand why students did not view the hint when they were offered it.

References

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of the 2004 conference on Human factors in computing systems - CHI 04*. doi:10.1145/985692.985741

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. doi:10.2307/1175554

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications, 40*(11), 4715–4729. doi:10.1016/j.eswa.2013.02.007

Davies, R., Nyland, R., Bodily, R., Chapman, J., Jones, B., & Young, J. (2017). Designing technology-enabled instruction to utilize learning analytics. *Tech Trends, 61*(2). 155–161. doi:10.1007/s11528-016-0131-7

Dennen, V. P., Darabi, A. A., & Smith, L. J. (2007). Instructor–learner interaction in online Courses: The relative perceived importance of particular instructor actions on performance and satisfaction. *Distance Education, 28*(1), 65–79. doi:10.1080/01587910701305319

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. doi:10.3102/0034654315581420

Martin, B., & Mitrovic, A. (2006). The effect of adapting feedback generality in ITS. *Lecture Notes in Computer Science Adaptive Hypermedia and Adaptive Web-Based Systems,* 192-202. doi:10.1007/11768012_21

Melis, E., & Andrès, E. (2005). Global feedback in ActiveMath. *Journal of Computers in Mathematics and Science Teaching, 24*(2), 192–220.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20. doi:10.24059/olj.v16i3.267

Siemens, G. (2011, August 5). Learning and academic analytics [Web log post]. Retrieved from http://www.learninganalytics.net/?p=131

Zhou, Y., Freedman, R., Glass, M., Michael, J. A., Rovick, A. A., & Evens, M. W. (1999). Delivering hints in a dialogue-based intelligent tutoring system. In *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI–99)* (pp. 128-134). Orlando, FL.

APPENDIX

Annotated Bibliography

The purpose of this annotated bibliography is to summarize several articles primarily related to one of two relevant topics: learning analytics and intelligent tutoring systems. The first section briefly introduces the concept of learning analytics, and the second and larger section discusses intelligent tutoring systems, as well as other relevant computer-based instruction, most of which employ learning analytics data in some fashion.

Literature Search Methodology

The sources for the annotated bibliography were found in a variety of ways. Some of them were found using the search keywords "learning analytics" and/or "intelligent tutoring systems" in several different research databases, including Scopus, ERIC, Google Scholar, and the general research search feature provided by the BYU Library website. Additionally, myself and two other BYU graduate students (Brice Colby and Tanya Gheen) who were working with similar research topics formed a research group on the article-hosting site Mendeley, wherein we all deposited articles that we found that we thought would be useful for each other. This was a source of several of the articles used in this work. Lastly, some of the sources came from a snowball-like effect as I examined the reference pages of sources that I had currently found in order to find additional relevant articles. The sources that are included here are ones that were determined to be relevant for reasons that will be discussed in the following two sections.

Introduction to Learning Analytics

This section primarily serves to briefly introduce and discuss learning analytics. These articles helped to deepen my understanding of what learning analytics data are and how they are used. One oft-cited definition of the term "learning analytics" can be found in the article by

Siemens (2011). The literature analysis conducted by Nunn, Avella, Kanai, and Kebritchi (2016) serves as a basic overview of how learning analytics data are used, as well as some of the common benefits and problems of this use. The article by Picciano (2012) offered suggestions on how learning analytics data should be used; he also warned about ways that such data could be potentially misused. Two additional articles (Davies, Nyland, Bodily, Chapman, Jones, & Young, 2017; Gašević, Dawson, & Siemens, 2014) further outline some of the issues currently facing those who use learning analytics data. Lastly, the article by Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García (2014) is present to provide a practical example of how learning analytics data can be used in situations that do not directly involve a system's design or performance (which is mostly what the following section is concerned with), but rather can be used to merely inform research. One common theme throughout many of these articles is that learning analytics data should be used to understand student behavior and modify instruction to meet their needs, which was one of the primary goals of our study.

Agudo-Peregrina, Á F., Iglesias-Pradas, S., Conde-González, M. Á, & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*, 542-550. doi:10.1016/j.chb.2013.05.031

By examining three different types of interactions (student-to-student, student-to-instructor, and student-to-content), the researchers who conducted this study set out to determine whether data about these three types of interactions could predict student success in online courses and face-to-face courses that use virtual learning environments

(VLEs), such as various learning management systems. The researchers found that that interaction data was only able to significantly predict student success in fully online classes they examined, but not in the face-to-face classes they examined. They additionally found that data about student-to-instructor interactions were the strongest predictors of student success.

Davies, R., Nyland, R., Bodily, R., Chapman, J., Jones, B., & Young, J. (2017). Designing technology-enabled instruction to utilize learning analytics. *TechTrends, 61*(2), 155–161.

In this article, the authors make several observations and offer commentary on how to improve learning analytics utilization. They begin by noting that there is much room for improvement for the use of data within education. One reason for this is because it can be difficult to design computer-based instruction that utilizes learning analytics data. Often, computer-based instruction is built merely with content delivery in mind, and it is not designed to collect or use data. In order to better make use of learning analytics data, the authors suggested that instructional designers should design computer-based instruction with data collection and use in mind. Designers should first identify what kind of data is necessary to understand the students' needs, and then design the system to collect that data. Once that data is captured, it can then be used by designers to improve the ability of computer-based instruction to remediate student misunderstandings and provide actionable information to students and teachers so that student learning can be improved.

Gašević, D., Dawson, S., & Siemens, G. (2014). Let’s not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71. doi:10.1007/s11528-014-0822-x

This paper highlights some of the problems facing those who use learning analytics. The primary dilemma that it discusses is that learning analytics are often used merely as a means of collecting and interpreting data which are then utilized to make changes to a system, without first analyzing that information through a lens of the principles and theories of instructional design and pedagogy. The authors noted that the effectiveness of the utilization of learning analytics data is often measured through performance on recall tests, such as quizzes, or through overall course grades. The authors expressed their worry that if learning analytics data are used primarily to boost test scores or overall grades, this could lead to situations such as when instructors “teach for the test,” which may further change the focus of education from learning to mere performance. The authors closed by again emphasizing the need for researchers, designers, and educators to use learning analytics data to enhance pedagogy and improve student learning, rather than to merely boost student performance in a way that is easily quantifiable, but which may not actually provide students with a deep understanding of the material.

Nunn, S., Avella, J. T., Kanai, T., & Kebritchi, M. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13-29. doi:10.24059/olj.v20i2.790

This literature review begins by pointing out that the purpose of learning analytics is to modify instruction in order to meet students’ needs. The authors contrast this with educational data mining, which tends to be more concerned with the acquisition of data itself in order to learn new information from that data. They then explained that the purpose of their literature review was to answer three questions: (a) How are learning analytics data used? (b) What benefits does the use of learning analytics data provide? (c) What are some of the difficulties associated with using learning analytics data? The authors found that learning analytics data are typically collected and analyzed in order to create a model that predicts student needs, which allows instruction to be altered to meet those needs. As this modified instruction is delivered, additional data is typically collected which can then be used to further refine the model, thus creating a cycle of data collection, analysis, model refinement, and instructional adaptation. The authors then identified several reported benefits that come from the use of learning analytics data. Some of these benefits include: allowing instructors to improve curriculum, allowing students to receive individualized instruction, and optimizing both student and instructor performance. Lastly, the authors noted some obstacles that users of learning analytics data face. One such difficulty is that the types of collectable data are still limited, and some data can be difficult to analyze in a meaningful way. Likewise, there are many ethical concerns around the collection and use of student data. The authors conclude by identifying some questions without clear answers, such as: Does personally identifiable data need to be collected and used? If so, how much control do students have over who can access and use their data? Do students need to provide consent for their data to be used if it is not personally identifiable?

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20. doi:10.24059/olj.v16i3.267

The author (Picciano) began this article by noting that college students are enrolling in online and blended classes in growing numbers. The presence of students in these online environments allows for a greater collection and analysis of data. Picciano pointed out that online companies often use data from their customers to model and predict behavior in order to be able to offer useful suggestions to customers of items they should purchase, or actions they should take. The author noted that educators are increasingly using student data in a similar way, and they argue that this should continue. Specifically, Picciano suggested that educators should use student data in order to inform educational decisions, designs, and practices. One specific use of student data that he proposed was using data about student performance in order to predict which students may be struggling and in need of an early intervention from the instructor to ensure that they successfully complete the course. He also noted that data could be used to evaluate the effectiveness of the current curriculum. Picciano closed his article by identifying a few concerns that those who seek to use student data may face. He noted that the largest issue that should be kept in mind is the potential dissemination and misuse of student data. In an age where data from all sorts of information repositories are often hacked, leaked, and abused, educators will need to take extra steps to ensure that their students' privacy is maintained.

Siemens, G. (2011, August 5). Learning and academic analytics [Web log post]. Retrieved from http://www.learninganalytics.net/?p=131

This brief article defines three terms: educational data mining, learning analytics, and academic analytics. Education data mining is a relatively new field that seeks to develop means to collect and study data produced in educational settings, in order to gain a deeper understand of students and their educational settings. Similarly, learning analytics involves the collection and analysis of student data and learning environment data, but in addition to merely understanding students and learning environments, the data is intended to be used to improve student learning and educational environments. Lastly, academic analytics is focused more on the collection, interpretation, and comparison of institutional data across individual institutions, regions, and nations.

Intelligent Tutoring Systems and Other Computer-Based Instruction

This section contains a seminal article written by Bloom (1984) wherein he identified what he calls “The 2 Sigma Problem.” The 2 Sigma Problem is repeatedly brought up in articles discussing intelligent tutoring systems (ITSs) as a problem that the developers and users of such systems are trying to overcome. An additional article (Kulik & Fletcher, 2016) addresses his work and is a meta-analysis of 50 different ITSs, wherein the authors proposed (based on their findings) that ITSs may be a feasible solution to this 2 Sigma Problem. The article by Chrysafiadi and Virvou (2013) is a literature review that details how various student models are used in the development of ITSs. Melis et al. (2001) described the design process for ActiveMath, a specific ITS that has been the subject of much research (including an additional article cited in this work).

An article by Keuning, Jeuring, & Heeren (2016) provides a literature review wherein the authors identified and defined several common types of feedback used by ITSs. Additionally, several articles (Melis & Andrès, 2005; Aleven & Koedinger, 2001; Martin & Mitrovic, 2006; Mitrovic & Martin, 2000; Zhou et. al, 1999) analyze various feedback systems within several specific ITSs.

The article by Baker, Corbett, Koedinger, & Wagner (2004) identifies “gaming the system” as a problem that occurs in some ITSs, and details a study that examines this behavior. The article by Chung et. al (2016) is included to detail a simpler alternative to ITSs proposed by the authors. Lastly, the article by Dennen, Darabi, and Smith (2007) is used mainly because it is a source that notes the increasing prevalence of online education (a statement which I noted in the journal-ready article portion of this report), but also because ITSs are often used within the context of online education, thus the suggested practices outlined in this article could be useful for online educators who makes use of ITSs.

Aleven, V., & Koedinger, K. R. (2001). Investigations into help seeking and learning with a cognitive tutor. In R. Luckin (Ed.), *Papers of the AIED-2001 workshop on help provision and help seeking in interactive learning environments* (pp. 47-58).

The researchers in this study utilized an intelligent tutoring system that provided hints of increasing specificity to geometry students when they requested help. The researchers wished to determine exactly how often students engaged in help-seeking behavior, as well as how much help they asked for when they asked for it. The researchers found that students requested help on 29% of the problems (on average), and that 81% of the time when students requested help, they kept requesting help until they reached the final hint (called the "bottom out hint"), which provides them with all the steps they need to perform to get the correct answer. The researchers speculated that this could mean that when students do not know the answer, they may be more interested in performance (simply providing the correct answer), rather than actual learning (understanding why those steps will provide the correct answer). One strength of the feedback system that we used in our study was that students had to make an additional attempt (requiring different input) to receive each successive hint. Additionally, our bottom out hints typically only provided the correct information for one specific component of the correct answer, thus (ideally) requiring students to do the work necessary to better learn the material.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of the 2004 conference on Human factors in computing systems - CHI 04*. doi:10.1145/985692.985741

In this study, researchers had a group of middle school students use a computer-based tutor to teach them various math skills. Students took a pre-test and a post-test to assess their abilities before and after using the tutor. The researchers covertly observed the students' behavior, and found that during their observations, 24% of students gamed the system at least once, and that 11% of the students spent more than 10% of their work time gaming the system. They found that the amount of time gaming the system had a negative correlation with their post-test score, even after controlling for the pre-test score, and overall academic achievement.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16. doi:10.3102/0013189x013006004

This paper began with Bloom citing some work done by two of his graduate students that had recently been published. These studies compared the results of students who were randomly assigned to either be in a “conventional” class of 30 individuals, or to have an individual tutor. Bloom then noted the findings from these studies: across grade levels and subjects, students who had the individual tutor scored on average two standard deviations higher than students in the control group. Bloom identified this as the titular “2 Sigma Problem.” Using this information, Bloom and his research team explored possible ways of trying to bridge this two-standard-deviation-gap (i.e., the 2 Sigma Problem) in more conventional classroom settings, most of which involved combining multiple methods, such as using mastery-based learning, encouraging teachers to engage their lower-achieving students, and having teachers focus on higher-level skills such as problem-solving and creativity.

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications, 40*(11), 4715–4729. doi:10.1016/j.eswa.2013.02.007

According to the authors, designers typically utilize some sort of student model when constructing computer-based tutors. The authors explained that a student model is a system that uses available data about a student to extrapolate what the student may know or not know about the topic being taught by the computerized tutoring system. This model allows the tutoring system to adapt instruction and/or learning activities (such as practice problems) in order to meet the student's individual needs. According to the authors, the elements that must be addressed when creating a student model include: (a) which characteristics of the student will be modelled, (b) how those characteristics will be modelled, and (c) how the student model will be used. The authors identify and describe several different student modeling approaches including: overlay, stereotypes, perturbation, machine learning techniques, cognitive theories, constraint-based model (which is the type of model that we used in our feedback system), fuzzy student modeling, Bayesian networks, and ontology-based student modeling.

Chung, G. K., Delacruz, G. C., Dionne, G. B., Baker, E. L., Lee, J. J., & Osmundson, E. (2016). *Towards individualized instruction with technology-enabled tools and methods: An exploratory study* (Rep. No. 854). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

While Intelligent Tutoring Systems (ITS) are typically effective, they can often be difficult to produce. The authors of this study wanted to determine whether there might be simpler means of providing effective personalized instruction. To do this, they prepared several brief instructional videos about a variety of topics covered in pre-algebra and algebra classes. Next, they administered pre-tests to middle school students in pre-algebra and algebra classes to determine each individual student's strengths and weaknesses. Then, students in the treatment group were exposed to 45 minutes of individualized instruction, wherein they could watch videos covering the topics that they were assessed to be the weakest in (based on the pre-test). These students were then administered a post-test. (Students in the control group did not receive any additional instruction between the pre-test and post-test). The researchers found that students in the treatment group outperformed those in the control group, indicating that this may be a simpler alternative to ITS (though they acknowledge that further research is needed).

Dennen, V. P., Darabi, A. A., & Smith, L. J. (2007). Instructor–learner interaction in online courses: The relative perceived importance of particular instructor actions on performance and satisfaction. *Distance Education, 28*(1), 65–79. doi:10.1080/01587910701305319

This article begins by noting that online education is becoming more common. In this study, the researchers reviewed the available and recent literature to create a list of 19 different practices that they considered important for online instructors to do in order to best help their students. They surveyed instructors and graduate students from two different universities and asked them to rank how important they felt that each of those items were for student success and satisfaction. They found that both students and instructors tended to rank the practices similarly, with a few subtle differences. One such difference was that, while both students and instructors tended to rank "prompt feedback" and "detailed feedback" highly, instructors thought that students would value detailed feedback more than prompt feedback, whereas the students found that prompt feedback was more valuable than detailed feedback. The researchers speculate that one reason for this could be that students are more concerned with their grade than how to improve their performance. They conclude this article by making three recommendations to instructors of online courses, based on their findings: 1) "[maintain] frequency of contact," 2) "[have] a regular presence in class discussion spaces," and 3) "[make] expectations clear to learners."

Keuning, H., Jeuring, J., & Heeren, B. (2016). Towards a systematic review of automated feedback generation for programming exercises. *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE 16,* 41-46. doi:10.1145/2899415.2899422

These authors conducted a small literature review of 69 studies that examined the feedback given by various computer-based tutoring systems that are used to teach students how to code. They identified four common types of feedback given by those systems. The first type, "knowledge about task constraints" (KTC), is when the system alerts the student that they are missing a specific required element from their code (such as the statement "FOR"). When students are offered additional instruction about subject matter that they may not fully understand, this is called "knowledge about concepts" (KC) feedback. "Knowledge about mistakes" (KM) feedback occurs when the system notifies the student that there is an error in their work. The final common type of feedback, “knowledge about how to proceed” (KH) is the type that the authors identify as the most helpful. It occurs when the system provides the students with hints to help them complete a specific step or task. However, few of the studied systems offered this type of feedback (they found that the most common type of feedback was KM). The authors suggested that if more feedback systems provided KH type feedback, that this would be of greater benefit to students. This finding supports the work that we did in our study, since one of the primary purposes of our feedback system was to provide our students with hints that would help them know how to proceed.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. doi:10.3102/0034654315581420

These researchers conducted a meta-analysis of 50 different evaluations of intelligent tutoring systems. They found that, in general, studies utilizing tutoring systems that are well implemented and robustly designed have effect sizes comparable to studies using human tutors. They also found that in many of the instances when tutoring systems were shown to be less effective, it was generally for one of two reasons, either (a) the tutoring system was poorly implemented (either because the instructors used the system as a replacement for classroom instruction rather than as a supplement to classroom instruction, or because the instructors were not trained in the use of tutoring systems), or (b) the control group used for comparison was poorly chosen (i.e., having other differences from the experimental group besides only whether or not the group received computer-based tutoring).

Martin, B., & Mitrovic, A. (2006). The effect of adapting feedback generality in ITS. *Lecture Notes in Computer Science Adaptive Hypermedia and Adaptive Web-Based Systems,* 192-202. doi:10.1007/11768012_21

These researchers conducted two different studies to determine whether using generalized feedback (rather than specific feedback) within an intelligent tutoring system (ITS) that used a constraint-based model would be beneficial to students. They used a currently existing ITS that normally gave very specific feedback. In their first study, students in the treatment group received only general feedback in certain instances (where the researchers determined, based on past research, that general feedback would suffice), whereas students in the control group always received specific feedback. Their first study produced mixed results, with the general feedback leading to an improvement in student performance in some cases, no improvement in others, and worse performance in still others. In their second study, students in the control group again received specific feedback, whereas students in the experimental group received general feedback at first, but were given the same specific feedback as the control group after two failed attempts at the problem in question. The researchers were surprised by the results in this study, where the treatment group actually seemed to do worse. However, they speculated that this may have been due to a flaw in their design: since the treatment students were unaware that feedback would progress from general to specific, they may have believed that they fixed the error causing the first (general) message, and looked for a new error that they may have committed which triggered the second (specific) message, thus being confused and not realizing that the same error was triggering both messages. In order to avoid a similar problem, our feedback system was designed so that as students progressed through three levels of increasingly specific hints, they were still able to see the text of the previous, more general hints, thus making it apparent that the same error was triggering these hints.

Melis, E., & Andrès, E. (2005). Global feedback in ActiveMath. *Journal of Computers in Mathematics and Science Teaching, 24*(2), 192–220.

This article describes the operation of the "suggestion mechanism" (an intelligent tutoring system) within ActiveMath, an online learning environment. The authors began by distinguishing between "local feedback" and "global feedback." Local feedback tends to be concerned with providing students with help on a specific problem, such as by giving them hints to arrive at the correct solution. Global feedback, on the other hand, is concerned with providing a wider understanding of the material, beyond just the application to one specific problem. In order to provide global feedback, the suggestion mechanism within ActiveMath goes through five distinct stages: (a) action capturing, (b) basic diagnosing, (c) diagnostic reasoning, (d) suggestion reasoning, and (e) suggestion rendering. During action capturing, the system logs several key student actions, such as when they start and stop reading an article, when they start and stop an assignment, their responses on an assignment, whether those responses are correct, and even eye movement. During basic diagnosing, the system analyzes the captured student data, and compares it with data in a currently existing database to provide several "diagnoses," or assessments, of that student's behavior. During diagnostic reasoning, the "diagnoses" from the previous stage are combined and analyzed by applying this data to the student model of the student in question. Based on this information, a suggestion for performance improvement is compiled in the suggestion reasoning stage. This suggestion is then presented to the student in the suggestion rendering stage.

Melis, E., Andres, E., Budenbender, J., Frischauf, A., Goduadze, G., Libbrecht, P., Pollet, M., & Ullrich, C. (2001). ActiveMath: A generic and adaptive web-based learning environment. *International Journal of Artificial Intelligence in Education (IJAIED), 12*(4), 385-407.

This article chronicles the design of ActiveMath, a web-based interactive learning system designed for a college-level algebra course. ActiveMath was designed to be adaptable according to the needs (and interests) of the student, in part by allowing students to freely navigate through the content, rather than having the content being delivered in a fixed, linear fashion. One feature of ActiveMath is the course generator, which receives information from the student as the student creates goals of what they want to learn. Based on the information, the system generates a custom course for the student to help them meet their learning goals.

Mitrovic, A., & Martin, B. (2000). Evaluating the effectiveness of feedback in SQL-Tutor. *Proceedings International Workshop on Advanced Learning Technologies. IWALT 2000.*

Advanced Learning Technology: Design and Development Issues. doi:10.1109/iwalt.2000.890591

The researchers in this study performed an evaluation of an Intelligent Tutoring System (ITS) that they designed to help university students learn SQL. The ITS used constraint-based modelling to assess the correctness of student responses. The ITS used six levels of feedback that went from broad to specific feedback based on how many incorrect attempts the students made, and how much help they requested. They found that students learned fastest and with the fewest errors when they were given feedback that provided hints to help them correct their errors, as opposed to broad feedback of simply whether or not their answers were correct, or specific feedback in the form of the correct answer.

Zhou, Y., Freedman, R., Glass, M., Michael, J. A., Rovick, A. A., & Evens, M. W. (1999). Delivering hints in a dialogue-based intelligent tutoring system. In *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI–99)* (pp. 128-134). Orlando, FL.

The researchers in this study sought to improve an intelligent tutoring system (ITS) called “CIRCSIM” which was used by medical students taking an anatomy class. To do this, the researchers interviewed human tutors that worked with students within that anatomy class to determine some common techniques for providing hints to students. From the interviews, they were able to categorize six different hinting strategies (specific to students in that class) which they integrated into their ITS. While their method of working with human tutors to identify the types of hints given to students is a good method that can be used by other ITS developers, it does not seem like the specific hint strategies that they identified would be generalizable or useful to contexts outside of that specific class or ITS.