



Analysis of Student Interactions with Browser-Based Interactive Simulations

Kyle Joe Branch, University of Utah

Kyle Branch is a second-year graduate student at the University of Utah Department of Chemical Engineering. He has helped develop and teach a freshman laboratory course using the methods described in this paper. His main research interest is in engineering education, focusing on the creation and analysis of interactive simulations for undergraduate chemical engineering courses.

Prof. Anthony Edward Butterfield, University of Utah

Anthony Butterfield is an Assistant Professor (Lecturing) in the Chemical Engineering Department of the University of Utah. He received his B. S. and Ph. D. from the University of Utah and a M. S. from the University of California, San Diego. His teaching responsibilities include the senior unit operations laboratory and freshman design laboratory. His research interests focus on undergraduate education, targeted drug delivery, photobioreactor design, and instrumentation.

Analysis of Student Interactions with Browser-Based Interactive Simulations

We have developed open-sourced interactive browser-based simulations that model realistic core engineering systems. Our simulations use JavaScript and HTML-5 to insure that the code is platform-agnostic and functional on all devices with a modern browser, avoiding some of the dissemination hurdles with educational Java applets or mobile apps. For each use of the simulations, we track student mouse movements and clicks, keyboard events, event times, screencast use, correlation with hands-on design project success, and more, leading to a large database that may be mined for pedagogical insights.

We have had remarkable success using these simulations while coupling them to collaborative, open-ended, hands-on design projects within the setting of a freshman design laboratory. In this course, students individually conduct experiments with the simulations before they come together as teams to design and build a process or product that relies on related core engineering theory.

Pre- and post-course surveys and tests were used to assess the teaching potential and students' evaluation of the simulations as course materials. Resulting student evaluations are far more positive than those found in a comparable engineering course using traditional pedagogy and static text-book assignments. Student learning was demonstrably improved along with student confidence in a variety of engineering skills. Our findings suggest that the simulations facilitate hands-on active and collaborative learning earlier in our students' academic career by making complicated engineering theory more accessible.

The resulting database of simulation usage data has been effective in detecting and responding to usage patterns of successful and unsuccessful students, allowing for iterative development of educational material. For example, ensemble averages of mouse location for successful and unsuccessful attempts in a spectrophotometer simulation revealed that unsuccessful students did not understand the need to properly calibrate. Student study habits and problem solving strategies also are evident in such data. Finally, we have found usage tracking data to be effective in improving user experience; for example, we detected attempts to interact with non-interactive elements of the simulation, prompting us to add interactive functionality to these elements.

By collecting real-time data on how student complete their homework, including both correct and incorrect attempts, we are able to both refocus our in-class discussions to address quantified weaknesses and add automated instructional supports in simulations to address errors at the moment they are detected. We believe, using such data, we will be able to bring some of the benefits of in-person active and collaborative learning to online simulations.

Introduction

It has been shown that online learning techniques may result in better student performance than face-to-face interactions¹. Additionally, research has shown that young students are able to learn material equally well from both hands-on projects and realistic

simulations of the same projects^{2,3}. No significant difference has been found in the quality of long-term knowledge of scientific concepts learned in these two ways⁴. It has also been shown that using interactive simulations to augment traditional lecture-based learning in an introductory physics course helped improve student understanding of key material and reduced their preconceived misconceptions⁵.

Educational data mining is an emerging area of research, focused on the computational analysis of educational data in order to gain insights on student learning⁶. Educational data mining on large volumes of logged student usage data has been used systematically to analyze student learning outcomes⁷. After analyzing the learning styles and habits of online users, automatic yet personalized supports can be offered to students; users have found this model to be effective⁸.

We have recently implemented a freshman-level chemical engineering laboratory course, which incorporates traditional lecture-based material, hands-on projects, and interactive simulations. Our interactive simulations collect detailed student usage data, supporting the use of educational data mining. Our post-class evaluations found that students enjoyed this teaching style and felt that they learned a great deal more, significantly preferring this course over a traditional lecture-based course covering much of the same material⁹. On pre- and post-tests, students performed better in every question, by 24% on average.

Methods

We have developed interactive simulations designed to realistically model laboratory experimentation¹⁰. These simulations cover a wide variety of engineering concepts, including heat, mass, and momentum transfer, electrical circuits, process control, thermodynamics, reaction kinetics, and microbial growth. Because these simulations enable students to see in real-time how the alteration of system parameters affects the rest of the system, they are valuable for aiding students in developing their engineering intuition.

Using these simulations, students are assigned to determine an unknown parameter in the system; in order to do so, they must alter properties and observe how the simulated system changes. For example, we model a spectrophotometer in one simulation, depicted in Figure 1, where students are able to alter the chemical species, concentration, light intensity, wavelength, and so on. Students are tasked with determining unknown reaction rate constants; in order to do so, they must calibrate the spectrophotometer, set reasonable starting concentrations, run the simulation for sufficient time, and then use the resulting data to determine the reaction rate constant. However, the steps required to successfully solve for an unknown property can often be accomplished in a variety of ways, similar to laboratory experimentation; the simulations are open-ended, allowing students to use whichever method they wish.

To help students better understand the simulations, we have developed screencasts with worked example problems. After submitting an answer, students are given immediate feedback on whether or not they got the problem correct. Additionally, the correct answer and a list of all of the system variables are displayed, allowing students to identify any errors they may have

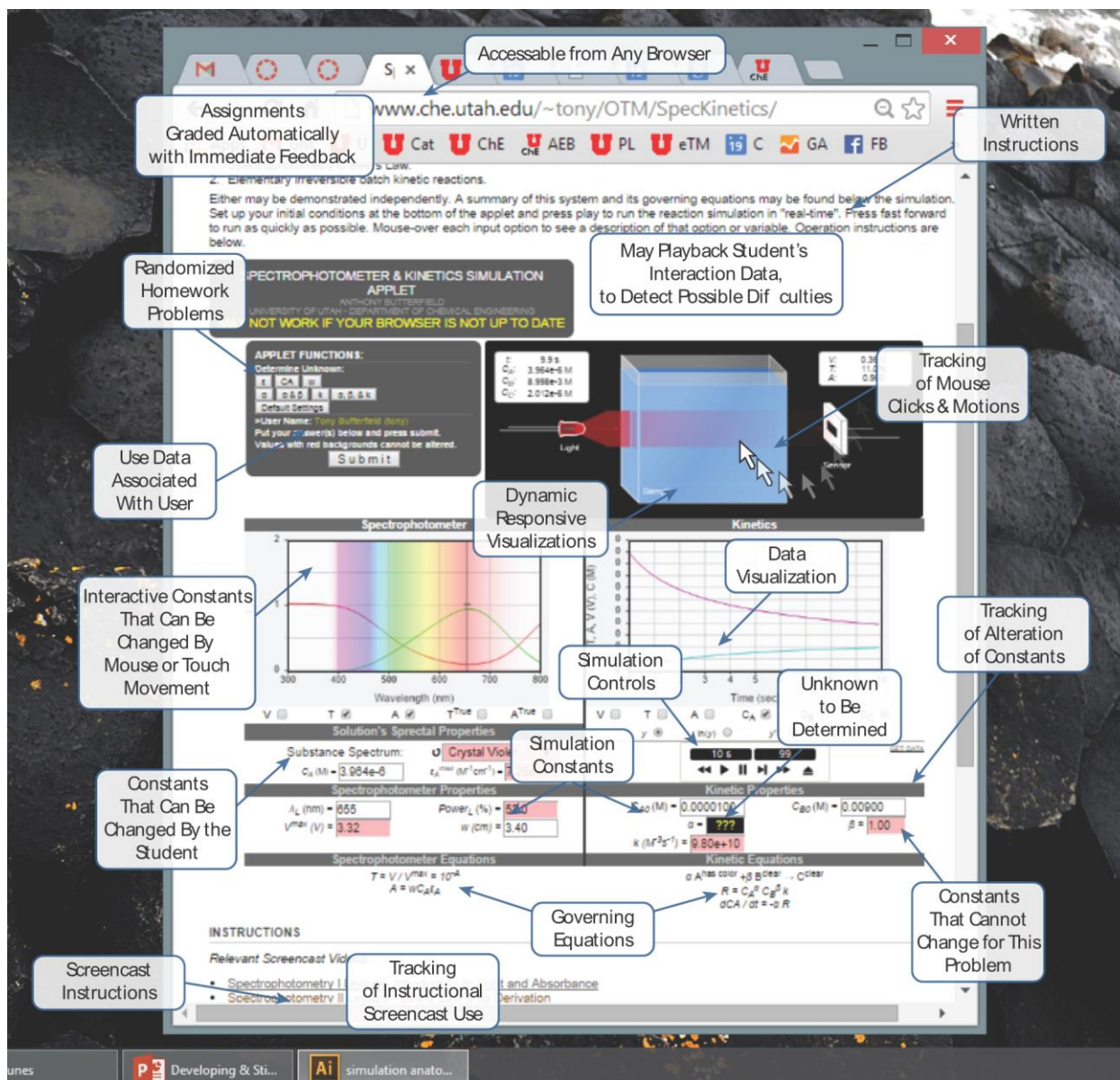


Figure 1 – A screenshot of our spectrophotometer simulation, with the components of the simulation labeled. Most of our interactive simulations have a similar structure.

made. The use of these simulations as homework problems helps prevent students from cheating, as each student is required to determine unique randomly generated constants.

The simulations are open-source and freely accessible to students and educators worldwide. In order to assure their ease of use, they have been designed to work on all major browsers, including Chrome, Firefox, Internet Explorer, and Safari, and devices including PC's, tablets, and smart phones. In order to eliminate potential security concerns, these simulations were made using HTML-5, JavaScript, and PHP; this eliminates the need for users to download, install, and run executable files.

Currently, our simulations are primarily used in our freshman and senior laboratory courses. In the freshman laboratory course, after students solve for a particular system parameter, they are asked to design and build a real-world version of the apparatus depicted in the simulation and then execute the experiment in the lab. The simulations allow students to develop ideas of how to run their experiment at their own pace; additionally, because students are prepared by their use of the simulated system, the project requires fewer lab resources and less class time.

Some of our simulations have been designed to allow students to input their real-world experimental data into the simulation and directly compare their results to theory. With this ability, we have tasked freshmen with estimating properties from their lab experiments, by manually altering simulation parameters until the simulation reasonably represents their data. This is an elementary alternative to more advanced techniques, such as nonlinear regression, to which freshmen have not been introduced. Seniors can also utilize this functionality to test the accuracy of the simulations by comparing analytical approximations to the simulated results. Additionally, senior students can determine empirical correction factors to enable the simulation to behave more realistically.

Our simulations also include the capability of tracking students' mouse movements, clicks, and typing, recording the time and location of each event in a secure MySQL database. This allows us to understand how the simulations are being used, and helps us determine characteristics of successful and unsuccessful student attempts. With this information, we can identify and respond to common misconceptions of students by adding simulation capabilities and improving instructions in the form of lectures and worked examples.

Results

We have used our simulation tracking capabilities for two years, recording the interactions of 177 students, totaling over 6,000 recorded student attempts. The first year's data includes records of whether or not the submission was correct, the time at which it was submitted, and the randomized simulation constants that were used. The simulations also record circumstantial information, including the student's browser, IP address, and operating system, which are primarily used for debugging the simulations. In the second year, we augmented our simulation tracking capabilities to also record all student interactions with the simulation, including the time and location of any mouse movements, clicks and field alterations. This data was collected for 103 of these students. Using this data, we can determine how long it takes students to complete different assignments, when they typically do their assignments, how many attempts they require, whether or not they are working on campus, and the number of and time between each click.

With this information, insights can be made looking at individual student attempts. Figure 2 shows a single student's a) incorrect, then b) correct attempt on our spectrophotometer simulation. During their first attempt, it is apparent that they were trying to alter simulation parameters, such as the width of the cuvette, by clicking and dragging on the diagram of the system, which was not a supported feature. In order to make the simulation more intuitive, we

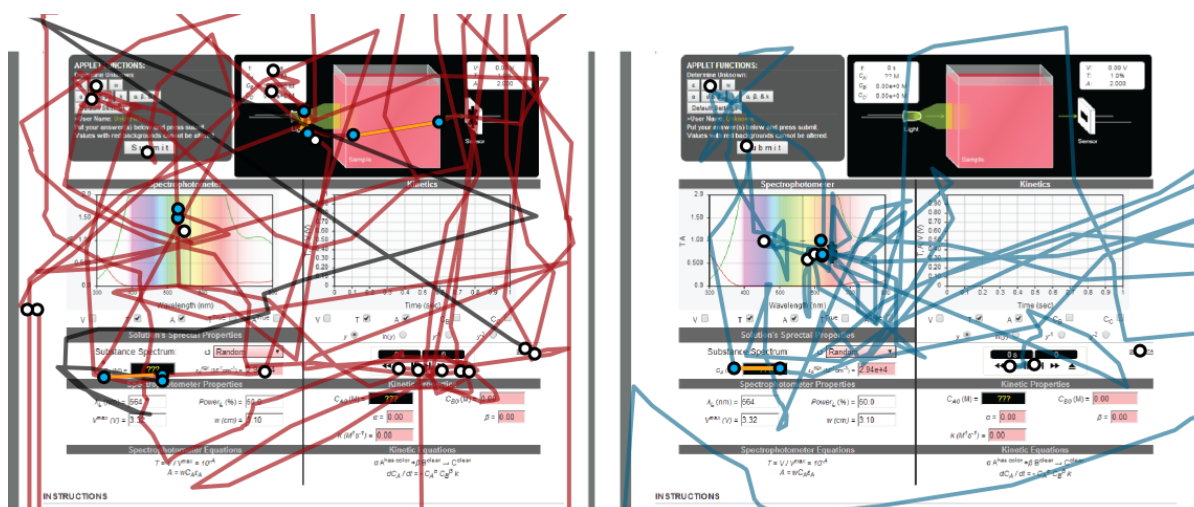


Figure 2 – Recorded mouse tracking data for an individual student's a) incorrect then b) correct attempt, while using our spectrophotometer simulation. It is apparent that on their first attempt, they were trying to alter simulation parameters, such as the width of the cuvette by clicking and dragging.

have now added this functionality. We have been primarily looking through ensemble averaged results and have only looked through a handful of individual student attempts; however we believe that there are many insights that could be made from investigating these attempts individually.

Using ensemble averaged mouse location data, we can see in which areas of the simulation students spend more time. Figure 3 shows a comparison of the mouse location distribution for correct and incorrect attempts on two different assignments in our freshman lab course using our spectrophotometer simulation. Lighter areas are locations in which students had

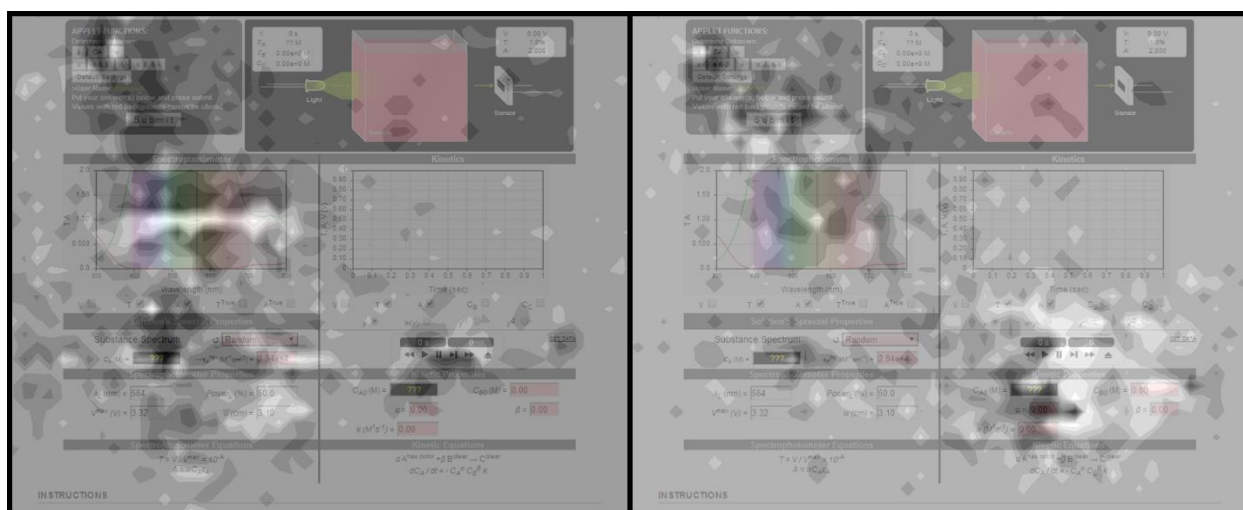


Figure 3 – Ensemble averaged mouse location data on our spectrophotometer simulation for students solving for a) maximum molar extinction coefficients, and b) reaction orders. Lighter areas indicate where mouse locations from correct submissions were more often, while darker areas indicate where mouse locations from incorrect submissions were more often. Data is on a truncated log-scale, where grey areas indicate that the percent of correct and incorrect submissions were very similar.

their mouse more of the time on correct submissions than on incorrect submissions. The contours are on a truncated log scale, where slight differences have been truncated to a neutral gray.

Figure 3 a) shows the mouse location data for students solving for a maximum molar extinction coefficient. In order to solve for this, students need to calibrate their

spectrophotometer. It can be seen that the incorrect submissions spend less time calibrating, shown by the white spot on the left. Also, it can be seen that incorrect solutions spend more time altering constants just below that; this set of constants is not required in determining the correct answer. On another problem, shown in Figure 3 b), students are tasked with determining a reaction order. For this, students are no longer required to calibrate their spectrophotometer, but instead they need to run the reaction. For this, students will primarily need to use the right half of the simulation. It is apparent that correct submissions' mouse locations are more often in this area of the simulation. Such results from our simulation data are used to iteratively improve lecture material and documentation on the simulation web site. In essence, with these tools we are able to detect and respond to student misconceptions in near-real-time.

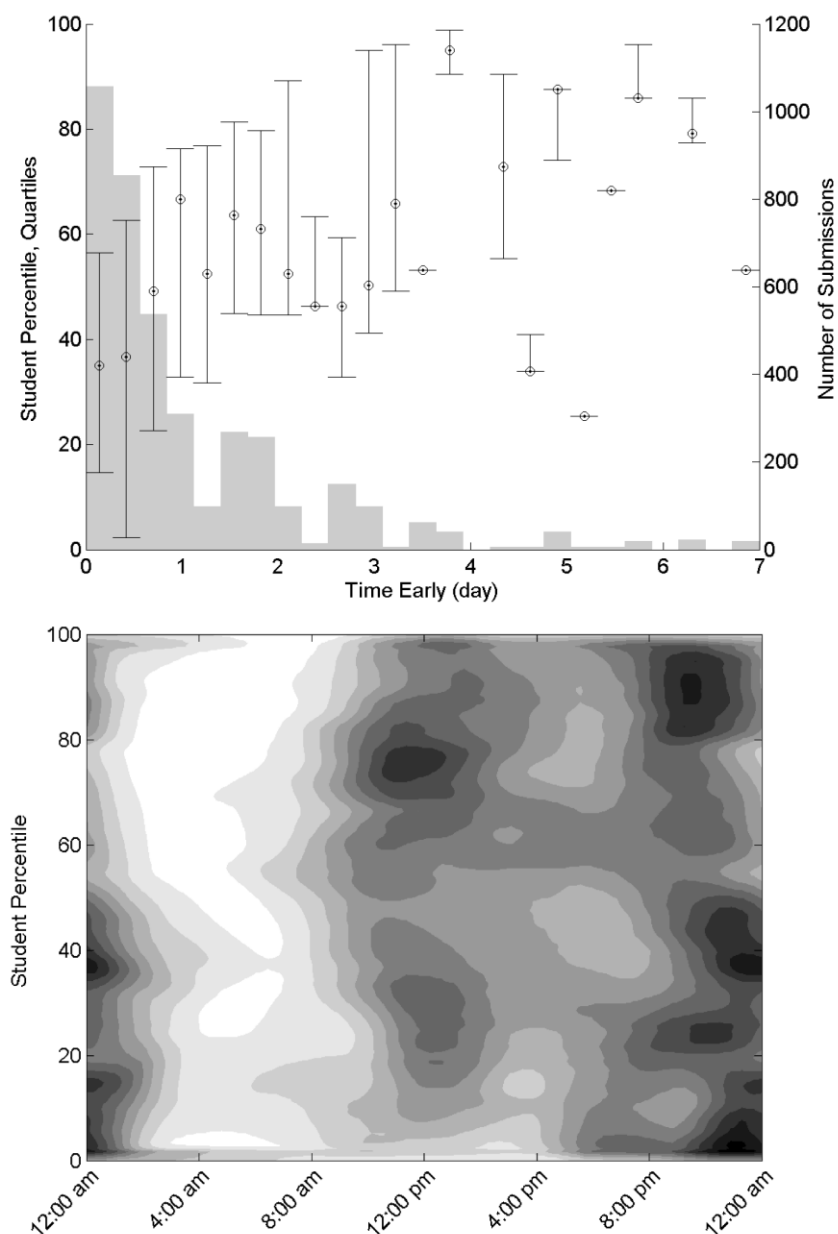


Figure 4 – a) The quartiles of student percentiles for all submissions in a particular before assignments are due. The grey bars shown indicate the total number of submissions in the time interval. b) Student percentile and the time of day they do their homework. It has been shown that higher percentile students do their homework earlier in the day, and don't stay up as late as their peers.

Our logged data may also be used to determine differences in students' study habits based on their performance in the course. Figure 4 a) shows that although not very many students start working on their assignments days before it is due; those who do tend to perform well in the class. We can also see what time of day students are typically working on their assignments, as show in Figure 4 b). With this, it can be seen that students primarily do their homework in the morning or evening. Higher percentile students seem to be more likely to work on their homework in the morning. Also, when they do their homework at night, they tend to do it earlier than low percentile students. High percentile students usually don't work on their homework after midnight, while it's not uncommon for others to work until 2:00 in the morning. This suggests that working late at night may be impeding student performance. This information may be presented to students to encourage healthier study habits and to illustrate the importance of time management; however, it is also possible that working late at night, while associated with poorer student performance, is not its direct cause.

With our usage data, we can also work to eliminate inauthentic problem-solving strategies. For example, to detect guessing, we monitored behaviors such as submitting an answer very quickly, not clicking very many times, and using the previous correct answer or the same answer twice in a row. In the future, if a student's attempts exhibits several of these behaviors, then their submissions will be flagged as guessing and they will not receive credit, in order to encourage legitimate problem solving. We also determined students who are guessing tend to do so late at night. Between 9:00 PM and 3:00 AM, 63% of the total guessed attempts occurred, whereas only 22% of submissions that were not characterized as guessing occurred during that period. Once again such findings may inform both faculty and students as to how best to use simulations for homework assignments.

Conclusions

Our simulation tracking capabilities have helped us identify the misconceptions and study habits of our students. By analyzing selected individual student attempts, we have improved the simulation user interface, adding functionality that students intuitively assumed to exist. We expect, with the tracking tools we have developed, that further insights could be achieved through comprehensive examination of these individual attempts.

We have visualized our tracking data using ensemble averages, showing differences in the mouse location distribution between correct and incorrect submissions. For example, Figure 3 a) shows that, for a particular assignment, students who submitted incorrect submissions spent less time in the calibration section of the simulation and more time altering constants that are not required for this problem. Such common misunderstanding are easily detected using our tools, whereas it would take TA's several hours to come to the same conclusions in hand-grading hundreds of problems. Such misunderstandings could be remediated with an increased emphasis on topics, such as calibration, during class discussions.

We can also use this data to understand trends in student behavior. Figure 4 a) shows that those who do their homework assignments earlier tend to be higher percentile students. Additionally, Figure 4 b) shows that higher percentile students are more likely to do their

homework in the morning, and lower percentile students tend to stay up later into the night working on them. Anecdotally, these trends are not surprising, but now we can clearly measure them.

In our future work, we plan to automate many of the reasonable instructor responses to such data. We may, for example, offer an automated reminder to calibrate if they have not in previous attempts, or we may have the site advise students to do their homework earlier if they're waiting until the night before it's due. We are also adding capabilities to detect unit conversion errors, and to notify the instructor when a particular student is having an abnormal level of difficulty so that in-person interventions may occur. With the ability to analyze each student's problem solving strategies for each assignment, we anticipate substantial gains.

Bibliography

- [1] Jones, K., "Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies," U.S. Department of Education, Office of Planning, Evaluation, and Policy Development Policy and Program Studies Service (2009).
- [2] D. Klahr, L. M. Triona, and C. Williams, "Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children," *J. Res. Sci. Teach.*, vol. 44, pp. 183–203, 2006.
- [3] L. M. Triona and D. Klahr, "Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments," *Cogn. Instr.*, vol. 21, no. 2, pp. 149–173, 2003.
- [4] M. D. Renken and N. Nunez, "Computer simulations and clear observations do not guarantee conceptual understanding," *Learn. Instr.*, vol. 23, pp. 10–23, 2013.
- [5] A. Jimoyiannis and V. Komis, "Computer simulations in physics teaching and learning: a case study on students' understanding of trajectory motion," *Comput. Educ.*, vol. 36, pp. 183–204, 2001.
- [6] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *Syst. Man, Cybern. C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, 2010.
- [7] W.-Y. Hwang, C.-B. Chang, and G.-J. Chen, "The relationship of learning traits, motivation and performance-learning response dynamics," *Comput. Educ.*, vol. 42, no. 3, pp. 267–287, Apr. 2004.
- [8] L. Zhang, X. Liu, and X. Liu, "Personalized Instructing Recommendation System Based on Web Mining," *9th Int. Conf. Young Comput. Sci.*, pp. 2517–2521, Nov. 2008.

- [9] A. Butterfield, K. Branch, E. Trujillo, "First-Year Hands-On Design Course: Implementation & Reception", *Chem. Eng. Ed.*, in press.
- [10] A. Butterfield, "Online Teaching Material," 2011. [Online]. Available: <http://www.che.utah.edu/~tony/OTM/>. [Accessed: 31-Jan-2015].