An analysis of peer-submitted and peer-reviewed answer rationales, in an asynchronous Peer Instruction based learning environment

Sameer Bhatnagar
Polytechnique Montreal

Michel Desmarais
Polytechnique Montreal

Chris Whittaker
Dawson College

Nathaniel Lasry
John Abbott College

Michael Dugdale
John Abbott College

Elizabeth S. Charles
Dawson College

ABSTRACT

This paper reports on an analyis of data from a novel *Peer Instruction* application, named DALITE. The Peer Instruction paradigm is well suited to take advantage of peer-input in web-based learning environments. DALITE implements an asynchronous instantiation of peer instruction: after submitting their answer to a multiple-choice question, students are asked to write a rationale for their choice. Then, they can compare their answer to other students' answers, and are asked to choose the best peer-submitted rationale among those displayed. We engaged in an analysis of student behaviour and learning outcomes in the DALITE learning environment. Specifically, we focus our investigation on the relationship between student proficiency, how students change their answers after reading each others' writings, and the peer-votes they earn in DALITE. Key results include i) peer-votes earned is a significant predictors of success in the course; ii) there are no significant differences between strong and weak students in how often they switch from the correct answer to a wrong answer after consulting peer-rationales, or vice versa; iii) even though males outscore females in conceptual physics questions, females earn as many votes from their peers as males do for the content they produce when justifying their answer choices.

Keywords

peer instruction, exploratory data analysis

1. INTRODUCTION

Active learning encompasses a broad movement in modern pedagogical practices, including any activities which engage the student as a part of the learning process, instead of passively receiving information during a traditional lecture. Such activities should encourage the student to read, write and discuss classroom content, as well as engage in higher order thinking tasks, such as synthesis and evaluation [1]. Active, cooperative, and collaborative learning practices have been shown to yield greater learning gains in science in engineering [8]. With the growing presence of on-line learning through instructional videos and accompanying readings, there is place for web-based activities which promote the same higher-order learning processes as those being used in more active classrooms.
This is where our research group found the need to develop the Distributed Active Learning Technology Integrated Environment (DALITE). The teacher-researchers in our group wanted a web-based homework system which would go beyond simply asking students for the answers to conceptual questions, by asking them to express the reasoning behind their thinking. This learning environment was meant to capture some of the higher-order thinking processes students engage in when reasoning about new concepts. DALITE is a system that would provide data on the mechanism of conceptual change, through the writings of students, as well as their evaluation of each other's work. What has emerged is an open source system which is being used in classrooms by learning science researchers who are also teachers.

Thus far, it has produced a dataset which can reveal new insights from the data on student production and consultation of answer rationales. Previous analysis of our work has already shown that students who use DALITE in college level physics classrooms do as well as those who use other on-line homework environments [2]. In the current study we analyze how the data on the production of rationales and the voting patterns can yield novel indicators of success and other characteristics of students.

This paper will begin with a description of the related field of Peer Instruction. The DALITE platform will then be described, as well as the most recent dataset collected. The focus of the analysis and results will be on the relationship between student proficiency, how students change their answers after reading each others' writings, and how many votes they earn for what they write. Finally we will discuss the potential and challenges that lie ahead, especially as student models are integrated into the DALITE system.

2. RELATED WORK

2.1 Peer Instruction

Peer instruction is a classroom practice popularized by Eric Mazur of Harvard University [3]. In its most common instantiation, the classroom script goes as follows:

1. The teacher displays a multiple choice question to the whole class, and asks everyone to reflect, and individually choose what they think is the correct answer. This is typically done by giving each student a handheld clicker, which transmits the answer to a receiver plugged into the teacher's computer.
2. The teacher displays a bar chart showing the distributions of answer choices for the whole class. The students are then prompted to discuss their answer choice with their peers for several minutes, after which they are given the opportunity to answer the question again using their clicker.
3. The teacher shows the new distribution of answers. Typically, after the peer discussion, there is a major shift towards the correct answer.

Making this a regular practice in class has been shown to yield higher learning gains [7] and lower dropout rates [4] compared to conventional, teacher-centered, lecture style courses. However it is very difficult to capture what is actually happening during the student discussions. What is actually being said to convince someone to change their answer (or at least change their rationale for their answer choice)? How does that relate to cognitive theories of learning? DALITE collects information exchanged in written form through Peer Instruction features embedded within a web based learning environment, namely answer rationales and votes. The information hereby collected allows us to better address the above questions empirically.

3. THE DALITE PLATFORM

DALITE is a web-based drill and practice platform that contains introductory level physics problems. It has an interface for the student to work on physics problems, and a teacher interface to manage the learning content.

3.1 Student interface

Students log into DALITE, and work on an assignment which typically contains four to six multiple choice questions. For each question, there are three screens they must flip through, each with the following structure:

1. The question is displayed, and the student selects one of the multiple choice answers. They are then prompted to write a couple of sentences that explain why they selected their answer choice. These little paragraphs will from now on be referred to as "rationales".
2. Once a rationale is given, the system presents two columns: one for their answer choice, and one for another choice to the question. Each column contains four rationales, written by previous students. The aim is to give students a chance to reflect on their thinking by providing them with an opportunity to compare and contrast other rationales and change their mind. The student is prompted to read the rationales from the two columns, and decide whether they would like to keep their choice, or switch. What's more, the student is asked to choose one rationale out of the ones displayed that they best like. They can also simply cast an "empty ballot", in effect saying that none of the other students' rationales were convincing. This up-voting process is anonymous.
3. The third screen recaps everything that just happened: the question is shown, alongside their two answer choices (one from each of the previous two screens). What's more, the rationale they originally wrote is reflected back to them, right next to a rationale written by an expert for the correct answer.

3.2 Teacher Interface

When teachers login to the system, they can:

- upload new questions to the database. This requires that the question be of multiple choice format. The teacher must specify the correct answer, with a rationale justifying that answer choice. The teacher must also identify a "second best answer", which would be used for the second column of the second screen (described above) should the student answer correctly on their first attempt. Teachers can also add "tags" to the question, which describe the content of the question.
- build new assignments based on questions already in the system.
- observe the results of assignments done by their students. The current reporting tool gives the teacher a mini grade-book for each assignment, where each student is a row, and each question is described by two columns: one for the student's first answer, and one for their second answer. Teachers can quickly get a sense of where the students are getting confused, as cells are coded green for the correct answer, and red for the incorrect answer. Transitions from red to green are signs that the rationales in the database are doing their job of convincing students to move away from the wrong answer, while transitions from green to red show that the students' conceptual understanding is shallow.

4. THE DATASET

Although DALITE has been in use for the last five years, it was during the Fall semester of 2013 that a comprehensive dataset was collected in a systematic manner over the entire term. The cohort was comprised of 144 students, spread out in five groups, taught by four different teachers, across three colleges. The system was used to teach freshman year, calculus -based Newtonian Mechanics. This is at a level equivalent to grade 12 in high school in the US and other Canadian provinces.

4.1 Data from within DALITE

Over the course of the semester, 80 question items were assigned by the different teachers, 40 of which were completed by at least half of the entire cohort, providing data on over 7000 student-item pairs.

Each student-item pair in the dataset includes the initial answer, the rationale, and the final answer. A separate table in the database keeps a count of how many peer-votes are earned by any given rationale.

4.2 Data from classrooms

For each student in the five experimental groups, as well as one control group (which did not use DALITE), the following data was collected inside their classrooms over the course of the semester:

Pre-Post FCI The Force Concept Inventory (FCI)[5], is a questionnaire of 30 conceptual questions about the Newtonian concept of force. The exact same questionnaire was administered on the first day of class, and then again on the last day of class, for each of the groups, in order to compare the learning gain between the DALITE users and students who did not use DALITE. The item-by-item results of this questionnaire can be compared to a FCI dataset which holds the results of more than 13000 students from across Canada and the U.S.

Midterm & Final Exam Grades The Newtonian Mechanics course commonly has three major themes: Kinematics, Dynamics, and Laws of Conservation. This lines up with the three midterms for which each student's grade is recorded. Finally, for each student, the final exam grade is broken down by the result on the multiple choice section (typically more conceptual questions, and hence more similar to DALITE), and the long-answer section (typically computations and problem-solving).

5. RESULTS

During the Fall 2013 study, four experimental groups were assigned DALITE specifically as homework for their students. Following are the key results:

Student Success How well students succeeded on DALITE questions had 0.50 and 0.60 correlations with their performance on the conceptual, multiple choice part of their final exam, and the post-semester FCI questionnaire, respectively. This provides some measure of the reliability of this relatively new homework system.
Also a linear model was fit to predict a student's final grade based on statistics from their DALITE account. The fraction of questions students answered correctly out of those they attempted, as well as the total number of votes they accumulated, were both significant predictors of their final grade in the course ($R^2 = 0.24$, $p<0.001$). This predictive power of DALITE emerges as early as after the first third of the course, meaning the teacher can get early indicators of which students are at risk for the midterm.
In a related line of questioning, the data was partitioned by gender of the students. Male students did significantly better than female counterparts in all measures of conceptual understanding from the classroom (pre-term FCI score, pre-post term gain on FCI, conceptual questions on final exam). This is in line with previous work looking into the gender gap in introductory physics [6]. This gap was found in the DALITE data as well, with males getting 20% more of the questions items right ($p<0.001$).

Patterns in how students change their answer choices Over the course of the semester, students who started with the right answer, only switched to the wrong one 1 out of 10 times. However, when they started with the wrong answer, they switched to the correct answer 3 out of 10 times after reading their peers' rationales. This gives some measure of overall quality of the rationales currently in the database: the rationales to the wrong answers are not highly persuasive, and there are at least some rationales for the correct answers which can convince students to change their minds when they are wrong.

Factors affecting answer change When the data was separated into quartiles for the final course grade, it was found that strong students were as likely as weaker students to switch from the right answer to the wrong answer. In addition, the converse was also true: weaker students were as capable of switching to the right answer when they got it wrong on their first attempt. There was some effect herein due to the teacher: the experimental groups that regularly discussed DALITE homework in class, were significantly more likely to change their answer when in DALITE. In the group that used DALITE purely as extra homework, answer switches were much less likely ($p<0.001$).This may indicate that the students who are reminded that the system is a valuable tool, are more engaged with the system, and take the time to more carefully read each others' rationales.
The well known gender gap mentioned, males outscoring females in conceptual physics questions, interestingly disappears if we measure correctness based on the second attempt: female students choose the wrong answer 20% more often on their first attempt, but after reading peer-written rationales, they identify the correct choice just as often as males.

Who amasses more peer votes? Students from the stronger half of the cohort earned, on average, more than two times as many votes as those from the bottom half. What's surprising is that this pattern holds true for the wrong answers as well: even when the strong students are wrong, they are twice as convincing as their weaker peers. This is especially relevant in light of the fact that 1/3 of all the votes cast over the term were for rationales to wrong answer choices. In parallel to this finding, when we looked only at rationales justifying the correct answer choice, it was found that weak students earned as many votes as their stronger colleagues. This seems to indicate that even if a student did not perform as well on tests, when they were right on a particular conceptual question, they were able to justify their understanding as well as stronger students.
The gender gap discussed earlier, was also lost when looking specifically at the voting data. Even though males achieve higher grades on conceptual questions, females of all strengths earn as many votes for their rationales as the males. This tends to indicate that females produce content justifying their understanding

that is as valued by their peers as rationales written by males.

6. DISCUSSION

The key results described above show the potential for DALITE to be an effective tool for teachers to probe their students' deeper understanding of concepts in physics, and identify students at risk of failing midterms and final exams. The data on how students change their answers based on the writings of their peers, and which rationales they vote for, may give teachers and researchers insight on what words can trigger conceptual change in different types of students. Finally, the data shows that students who may not perform as well on summative evaluations, are still able to produce valuable content when justifying their understanding.

7. FUTURE WORK

Future directions of research on this project include capturing not just which rationales got voted for, but who is casting the votes, and in what context. The goal is to explore what features in student written text have an impact on changing peer conceptions of scientific concepts. Do students learn from stronger students, or only those within their Vygotskian zone of proximal development [10].

Another important direction would include collaborative filtering techniques, which are traditionally applied to recommender systems, such as in the e-commerce setting, where a users-by-item ratings matrix is used to predict what items new users would most likely enjoy. Recently such techniques have been applied in the context of educational data mining, where the matrix is now student-by-item performance, and factorization leads to estimates of the probability of another student getting a new item correct [9]. With the ratings data collected, the system may be able to deliver individualized rationales to different learners with the same misconceptions to the same question item. What is most promising is how this open-source tool creates a venue for learning science researchers to ask questions regarding higher-order learning processes, such as evaluation and synthesis, and for the EDM community to test-drive different text mining techniques in a real classroom setting.

8. ACKNOWLEDGMENTS

The strength of the DALITE platform resides in the database of student rationales, so the students who have used this platform for learning must be thanked for providing this rich set of data. This work has been funded through the Programme d'Aide à la Recherche sur l'Éducation et l'Apprentissage(PARÉA), administered by the Ministère d'Éducation et Loisirs de Quebec.

9. ADDITIONAL AUTHORS

Kevin Lenton, Vanier College

10. REFERENCES

[1] C. C. Bonwell and J. A. Eison. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC, 1991.

[2] E. Charles-Woods, C. Whittaker, M. Dugdale, N. Lasry, K. Lenton, and S. Bhatnagar. Beyond and within classroom walls: Designing principled pedagogical tools for students and faculty uptake. In *Computer Supported Collaborative Learning (in press)*, 2015.

[3] C. H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9):970–977, 2001.

[4] A. P. Fagen, C. H. Crouch, and E. Mazur. Peer instruction: Results from a range of classrooms. *The Physics Teacher*, 40(4):206–209, 2002.

[5] D. Hestenes, M. Wells, and G. Swackhamer. Force concept inventory. *The physics teacher*, 30(3):141–158, 1992.

[6] L. E. Kost, S. J. Pollock, and N. D. Finkelstein. Characterizing the gender gap in introductory physics. *Physical Review Special Topics-Physics Education Research*, 5(1):010101, 2009.

[7] N. Lasry, E. Mazur, and J. Watkins. Peer instruction: From harvard to the two-year college. *American Journal of Physics*, 76(11):1066–1069, 2008.

[8] M. Prince. Does active learning work? a review of the research. *Journal of engineering education*, 93(3):223–231, 2004.

[9] N. Thai-Nghe, L. Drumond, T. Horváth, A. Krohn-Grimberghe, A. Nanopoulos, and L. Schmidt-Thieme. Factorization techniques for predicting student performance. *Educational Recommender Systems and Technologies: Practices and Challenges*, pages 129–153, 2011.

[10] L. Vygotsky. Interaction between learning and development. *Readings on the development of children*, pages 34–41, 1978.