122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Seattle
Making Value for Society
Paper ID #13544

PrairieLearn: Mastery-based Online Problem Solving with Adaptive Scoring and Recommendations Driven by Machine Learning

Prof. Matthew West, University of Illinois, Urbana-Champaign

Matthew West is an Associate Professor in the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign. Prior to joining Illinois he was on the faculties of the Department of Aeronautics and Astronautics at Stanford University and the Department of Mathematics at the University of California, Davis. Prof. West holds a Ph.D. in Control and Dynamical Systems from the California Institute of Technology and a B.Sc. in Pure and Applied Mathematics from the University of Western Australia. His research is in the field of scientific computing and numerical analysis, where he works on computational algorithms for simulating complex stochastic systems such as atmospheric aerosols and feedback control. Prof. West is the recipient of the NSF CAREER award and is a University of Illinois Distinguished Teacher-Scholar and College of Engineering Education Innovation Fellow.

Dr. Geoffrey L Herman, University of Illinois, Urbana-Champaign

Dr. Geoffrey L. Herman is a visiting assistant professor with the Illinois Foundry for Innovation in Engineering Education at the University of Illinois at Urbana-Champaign and a research assistant professor with the Department of Curriculum & Instruction. He earned his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign as a Mavis Future Faculty Fellow and conducted postdoctoral research with Ruth Streveler in the School of Engineering Education at Purdue University. His research interests include creating systems for sustainable improvement in engineering education, promoting intrinsic motivation in the classroom, conceptual change and development in engineering students, and change in faculty beliefs about teaching and learning. He serves as the webmaster for the ASEE Educational Research and Methods Division.

Prof. Craig Zilles, University of Illinois, Urbana-Champaign

Craig Zilles is an Associate Professor in the Computer Science department at the University of Illinois at Urbana-Champaign. His current research focuses on computer science education and computer architecture. His research has been recognized by two best paper awards from ASPLOS (2010 and 2013) and by selection for inclusion in the IEEE Micro Top Picks from the 2007 Computer Architecture Conferences. He received the IEEE Education Society's Mac Van Valkenburg Early Career Teaching Award in 2010, a (campus-wise) Illinois Student Senate Teaching Excellence award in 2013, the NSF CAREER award, and the Univerisity of Illinois College of Engineering's Rose Award and Everitt Award for Teaching Excellence. Prior to his work on education and computer architecture, he developed the first algorithm that allowed rendering arbitrary three-dimensional polygonal shapes for haptic interfaces (force-feedback human-computer interfaces). He holds 6 patents.

PrairieLearn: Mastery-based Online Problem Solving with Adaptive Scoring and Recommendations Driven by Machine Learning

1. Introduction

Online homework systems have exploded in use in large introductory STEM courses in recent years[10], due to benefits to both students and instructors, such as immediate feedback, integration with online content, and reduced grading workloads, although some concerns have also been expressed[5]. Many current-generation online homework systems follow a model that is very similar to traditional paper homework, in which a relatively small number of questions are assigned each week, and students must complete each problem exactly once. Online-only features often include question randomization, so students all receive slightly different versions of each question, and immediate online grading and feedback.

Educational research and learning theory show that small numbers of single-practice problems may not be the most effective learning strategy for students. Mastery learning theory[7] shows that different students require different amounts of practice to achieve proficiency in a given skill, and that all students require repeated practice[1]. Additionally, spaced-repetition theory[2] provides evidence that it is more effective to space out repeated practices of the same or similar items[6].

To incorporate both mastery learning and spaced-repetition concepts into online homeworks, we developed the PrairieLearn web-based homework system. This system simultaneously models both student ability and question difficulty and guides students by adaptively awarding students different numbers of points to each question (positive for correct answers, negative for incorrect answers). The objectives of this system are to: (1) enable students to practice solving randomized problem variants repeatedly until mastery, (2) incentivize students to repeat questions until mastery is achieved, and (3) provide immediate feedback about their current mastery level to the student.

The PrairieLearn system was deployed in the course *Introductory Dynamics* at the University of Illinois at Urbana-Champaign, as part of a broader reform of the course[4;13]. More recently, PrairieLearn has also served as a key part of a new Computerized Testing facility at the University of Illinois[14]. In Section 2 we describe PrairieLearn in detail, before presenting student feedback data in Section 3 and analyzing student behavior with PrairieLearn in Section 4. Finally, conclusions are presented in Section 5.

2. The PrairieLearn system

The interface PrairieLearn presents to a student for a homework is shown in Figure 1. This is implemented as an open-source Node.js server and a JavaScript web-app. For each homework,

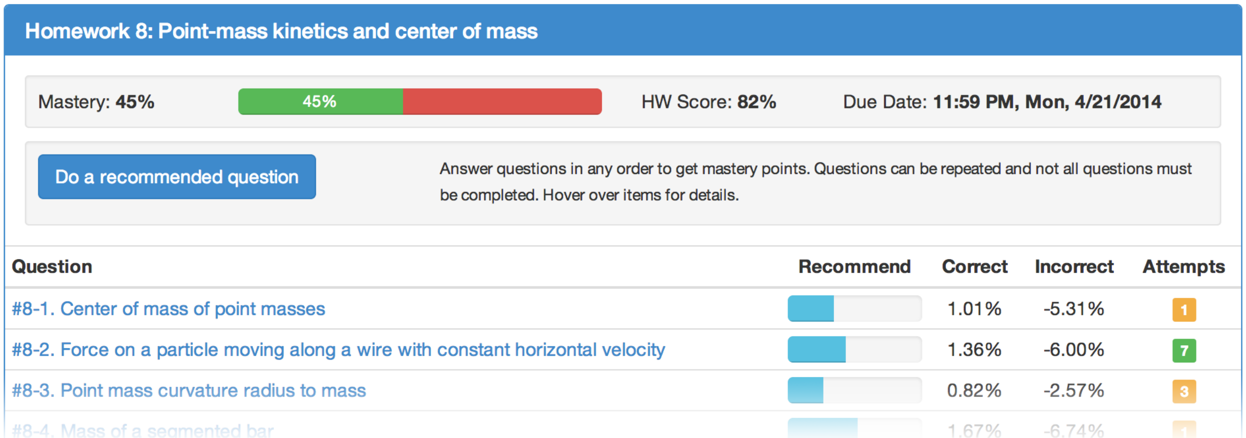

Figure 1: The PrairieLearn homework overview page, showing the mastery score and the list of questions, each with its current recommendation level, the number of points that will be awarded if it is answered correctly, the number of points that will be deducted if it is answered incorrectly, and the number of times that question has been previously attempted. Questions are listed in order by estimated difficulty. The "HW Score" is the score that the student will actually receive for this homework (a constant factor multiplied by the Mastery Score, capped at 100%). The "Do a recommended question" button will take the student to a randomly chosen question with a high recommendation rating, or they can click on a specific question to do it directly.

the student has a *mastery score* that reflects PrairieLearn's estimate of the student's ability on this homework assignment. To increase their mastery score, the student must answer questions correctly, in any order they choose. A student can attempt a question as many times as they like (whether answering correctly or incorrectly), but question parameters are randomized on every attempt so the answer will always be different. See Figure 2 for an example of a question within PrairieLearn.

To guide the student towards doing questions that will be productive and educationally valuable, PrairieLearn uses a model of the student's current ability (described below in Section 2.1) to adaptively show a recommendation rating for each question, and also to change the points awarded or deducted for correct or incorrect answers, respectively, to each question.

2.1. Student/question model

PrairieLearn uses a parametric model to describe the interaction between students and questions. In particular, it uses a three parameter logistic model[9], which describes student i by an ability parameter σ_i, and question j by a three-vector $\theta_j = (\alpha_j, \beta_j, \gamma_j)$. Then the *question response function* is given by:

$$\mathrm{Prob}(\text{student } i \text{ answers question } j \text{ correctly}) = f(\sigma_i, \theta_j) = \gamma_j + \frac{1 - \gamma_j}{1 + e^{\alpha_j(\beta_j - \sigma_i)}} \tag{1}$$

This function is shown graphically in Figure 3. While this model is common for many machine-learning tasks[9], it has also been introduced independently in the context of item

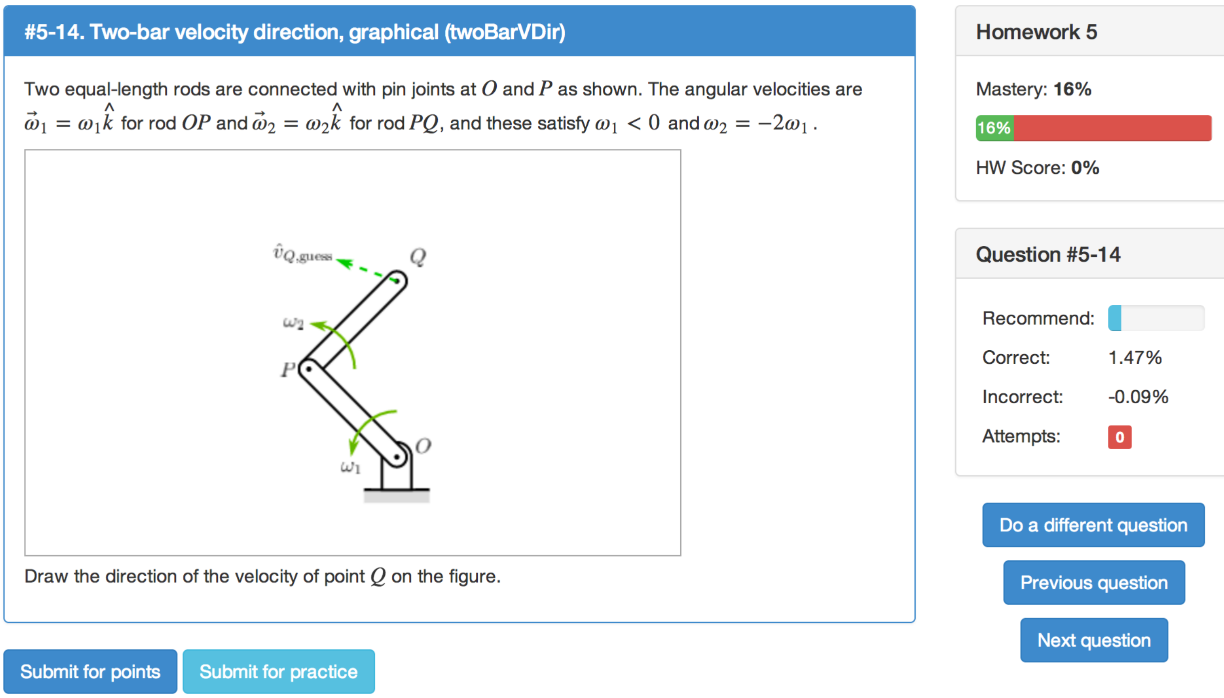

Figure 2: One question within PrairieLearn. The main box (upper left) contains the question itself, which in this case asks the student to draw the vector $\hat{v}_{Q,\text{guess}}$ as a guess to the true velocity direction vector $\hat{v}_Q$ of point Q. The question will be graded as correct if the true velocity direction is within a certain tolerance δ, that is, if $\|\hat{v}_Q - \hat{v}_{Q,\text{guess}}\| < \delta$. Other question formats within PrairieLearn include numerical answers and multiple choice. On the right side of the question page, the student can see their current mastery score, as well as question-specific information including the current recommendation level, the number of points that will be awarded if this question is answered correctly, the number of points that will be deducted if this question is answered incorrectly, and the number of times this question has already been attempted. The buttons in the lower left will take the student to the next or previous question (in order within the homework), or to a randomly chosen question with high recommendation rating (the "Do a different question" button). The buttons in the lower left allow the student to submit their current answer to this question either for points (positive or negative, depending on correctness), or for practice, which will grade the question but not update the student's score in any way. Submitting a question in either way will show the student whether they were correct, and what the true answer is, before returning them to a new random instance of this same question.

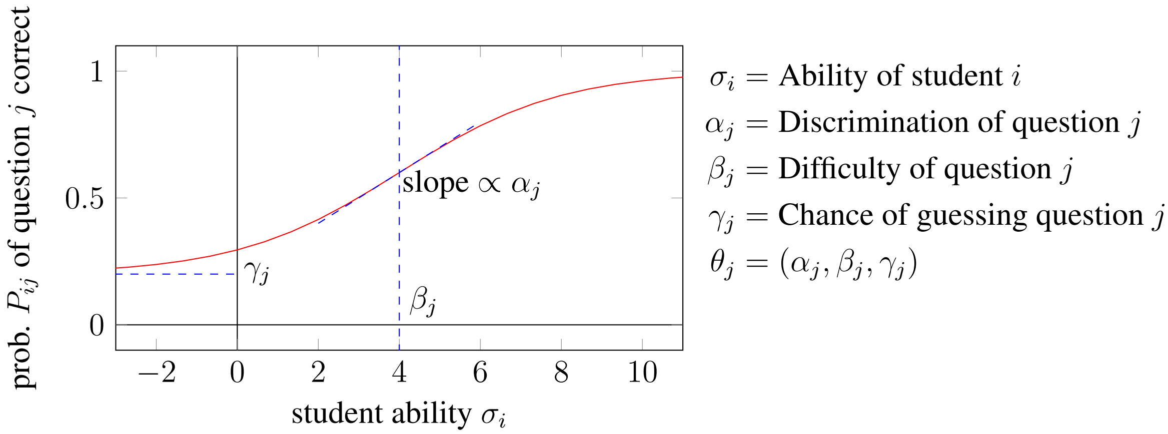

Figure 3: The logistic model that describes student/question interations within PrairieLearn.

response theory[11] within education, which is commonly used to understand student performance on assessments.

The Mastery Score m computed by PrairieLearn for student i is the average probability of that student answering a randomly-chosen question correctly. That is,

$$m = \frac{1}{M} \sum_{j=1}^{M} E[f(\sigma_i, \theta_j)]. \tag{2}$$

The model parameters in PrairieLearn are determined and updated in two separate phases. First, question parameters $\theta_j = (\alpha_j, \beta_j, \gamma_j)$ are initially chosen by the instructor based on professional judgment. Student parameters σ_i are initially all set to have a prior probability distribution given by a standard normal distribution $N(0, 1)$. As student i answers homework questions, their student ability distribution function is updated by a Bayesian update rule[8]. At this point, an "aging" term is included in the model to devalue repeated attempts at a question within a certain window, thereby encouraging spaced repetition by students. Second, after the homework is complete, question parameters θ_j are updated using maximum likelihood estimation (MLE)[3;8].

2.2. Adaptive scoring and recommendations

The benefit that student i will derive from attempting question j is modeled by a *utility function* $U(\sigma_i, \theta_j)$. While determining the true utility function would be very difficult in practice, PrairieLearn assumes that U is proportional to the derivative of the question response function f with respect to the student ability, so

$$U(\sigma_i, \theta_j) = \frac{\partial}{\partial \sigma_i} f(\sigma_i, \theta_j). \tag{3}$$

This choice reflects the insight that effective learning occurs when performing tasks for which there is some moderate chance of success, but when success is not always guaranteed[6]. The per-student adaptive question recommendations that PrairieLearn shows each student are proportional to the expected value $E[U]$, normalized to a standard scale across all questions.

The adaptive scores that PrairieLearn presents to students are simply the amount that their Mastery Score m will change if they get the question correct or incorrect. Assuming first correct and then incorrect answers, the Bayesian update rule described above can be used to compute the resulting updated student parameter distribution, and from this the resulting updated Mastery Scores m^+ and m^- can be determined, respectively, giving:

$$\text{points gained for correct answer} = m^+ - m \tag{4a}$$
$$\text{points lost for incorrect answer} = m - m^-. \tag{4b}$$

3. Student perceptions survey

A survey addressing student perceptions of the PrairieLearn system was administered in *Introductory Dynamics* in Spring 2014, with results as shown in Figure 4. That semester was the first semester in which PrairieLearn was used for credit and, to enable a side-by-side comparison, *Introductory Dynamics* in that semester used both PrairieLearn as well as the pre-existing non-adaptive homework system.

The survey results show that students like the idea of online homework (80% like against 4% dislike) and that they prefer it to written homework (77% preferring online against 10% preferring written). They generally work on the online homework themselves (63% agreeing versus 17% disagreeing) and they broadly find the online homework in *Introductory Dynamics* to be of the correct difficulty level (37% too hard, 42% just right, and 21% too easy).

When comparing the adaptive PrairieLearn system to the existing non-adaptive online homework system in *Introductory Dynamics*, students were asked to compare on two different dimension: (1) whether each system helped prepare them for exams, and (2) whether each system was helpful in learning that material. PrairieLearn was thought to be beneficial both for exam preparation (76% beneficial versus 13% non-beneficial) and understanding course material (63% beneficial versus 23% non-beneficial). In contrast, the pre-existing non-adaptive homework system was thought to be beneficial for understanding course material (61% beneficial versus 14% non-beneficial), but was not thought to be very helpful for exam preparation (33% beneficial versus 38% non-beneficial). Interestingly, PrairieLearn elicited significantly more extreme opinions about whether it was beneficial for understanding the material (30% very beneficial versus 12% very non-beneficial), when compared to the non-adaptive system (19% very beneficial versus 4% very non-beneficial).

In summary, the adaptive PrairieLearn system was considered by students to be more beneficial than the pre-existing non-adaptive system, but this effect was more pronounced for exam preparation than understanding of course material.

4. Student interactions with PrairieLearn

To understand how students interacted with the PrairieLearn system, data was collected during the Fall 2014 semester of the course *Introductory Dynamics* at the University of Illinois, with $N = 194$ students. In the following subsections, we consider different analyses of this data to understand whether PrairieLearn led to behaviors known to improve mastery, especially spacing and repetition[6].

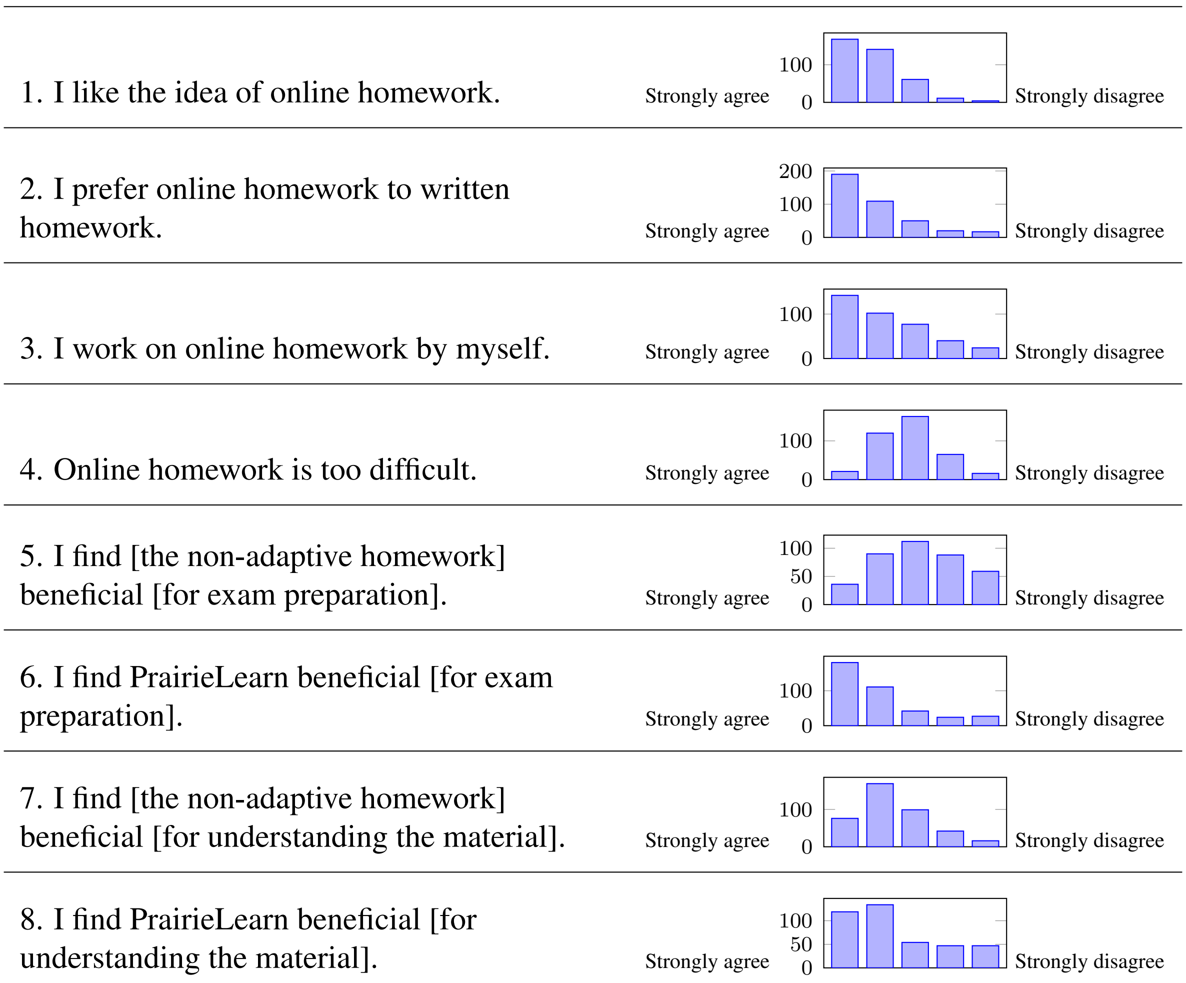

Figure 4: Student survey results ($N = 408$ students) from *Introductory Dynamics* in Spring 2014, when the adaptive PrairieLearn system was introduced in conjunction with the previous non-adaptive online homework system. See Section 3 for discussion.

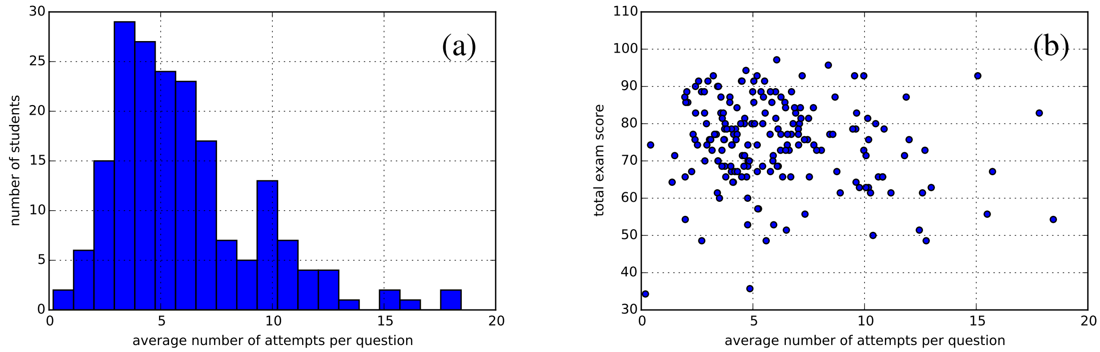

Figure 5: (a) Distribution of the average number of attempts per question (mean 6.0, standard deviation 3.2). (b) Correlation of students' exam score with the average number of attempts per question (correlation coefficient of -0.14). See Section 4.1 for discussion.

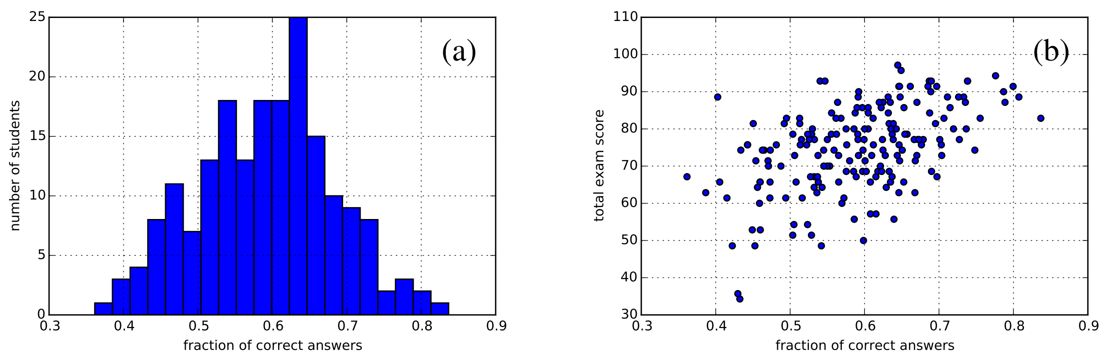

Figure 6: (a) Distribution of the fraction of correct answers (mean 0.59, standard deviation 0.09). (b) Correlation of students' exam score with the fraction of correct answers (correlation coefficient of 0.51). See Section 4.1 for discussion.

4.1. Per-student metrics

Figure 5(a) shows the distribution of the per-student average number of attempts per question, and Figure 5(b) shows the correlation of this quantity with the students' total exam score (two midterm exams plus one final). We see that students are repeating questions many times (mean of 6.0 attempts per question), which is encouraging from a repetition point of view, and that the number of questions solved by students does not correlate with exam scores (it is actually slightly negatively correlated).

The fraction of correct answers per student is shown as a distribution in Figure 6(a), and correlated against exam scores in Figure 6(b). The average student solves questions correctly 59% of the time, but there is wide variation. Unlike the number of attempts per question, the fraction of correct answers does correlate with exam scores (correlation coefficient of 0.51).

To understand how the fraction of correct answers and the number of question attempts interact,

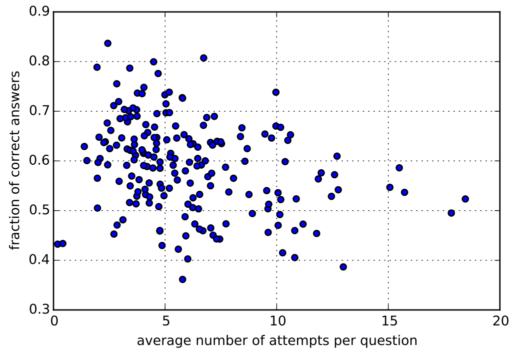

Figure 7: Correlation of the average number of attempts per question with the fraction of correct answers (correlation coefficient of -0.29). See Section 4.1 for discussion.

they are plotted against each other in Figure 7, which shows they are negatively correlated (correlation coefficient of -0.29). Combined with Figures 5 and 6, this suggests that some weaker students are indeed practicing more on average than stronger students, but not enough to entirely reach the same mastery level on average.

4.2. Per-question, per-student metrics

To understand how often students try to solve a particular question before their first successful attempt, Figure 8 shows several different histograms of the number of attempts per student-question pair. That is, for every student i and every question j, a number n_{ij} is computed and the histogram of n_{ij} is plotted. The subfigures within Figure 8 constructs n_{ij} in different ways.

Figure 8(a) shows the histogram of the number of attempts per question (mean 7.9) for questions with at least one attempt. This is somewhat higher on average than the mean of Figure 5(a), which is computed per-student, because students who make many attempts have less impact in Figure 5(a).

For each student i and question j, we say that question j is *never solved* by student i if that student attempts the question at least once, but never submits a correct answer to that question. Figure 8 shows the histogram of the number of attempts made for such questions. From this we see that students typically give up fairly quickly if they can't successfully solve a question, with very few students persisting beyond three incorrect attempts.

In contrast to never-solved questions, we say that question j for student i is *eventually solved* if the student submits at least one correct solution to that question. Figure 8(c) shows the histogram of the number of attempts at eventually-solved questions, which has a slightly higher mean (9.0 attempts) than all questions, as the never-solved questions have a much lower mean.

Figure 8(d) shows the number of incorrect attempts made by students on questions that they will eventually solve. In many cases students have zero incorrect initial attempts, meaning that their first attempt is successful, and we see that they typically solve the question correctly within the first few attempts, with very few questions having beyond three incorrect initial attempts. This is

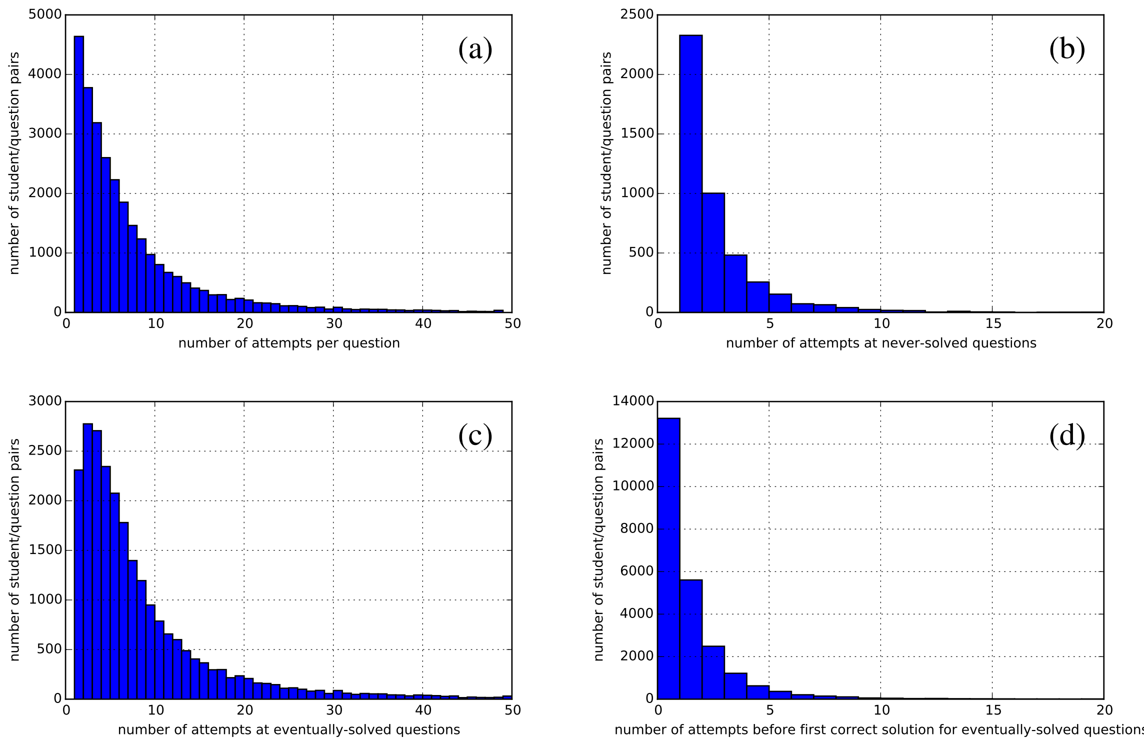

Figure 8: Attempt patterns per student-question pair. (a) Distribution of the number of attempts per question (mean 7.9, standard deviation 11.4). (b) Distribution of the number of attempts per question for questions that the student never correctly solves (mean 2.3, standard deviation 2.6). (c) Distribution of the number of attempts at questions that the student does solve correctly (mean 9.0, standard deviation 12.1). (d) Distribution of the number of attempts before the first correct solution, for questions that the student does solve correctly (mean 1.0, standard deviation 2.2). See Section 4.2 for discussion.

consistent with the observation from Figure 8(b) that students will typically not persist beyond about three incorrect attempts before giving up on a question.

4.3. Ordering of question attempts

As well as simply encouraging question repetition for improved learning, another objective of PrairieLearn was to encourage spaced repetition by the use of a devaluing term for repeated recent solutions of each question during the Bayesian parameter update (see Section 2.1). To see whether students are indeed varying the order of questions that they solve, Figure 9 shows question “trajectories” chosen by students for a single representative homework (Homework 4, on tangential/normal bases). For Figures 9(a)–(c), we take three individual representative students. For each student, we chronologically order all of their question attempts (the “attempt number”) and plot against this the question number of each attempt. Recall from Section 2 that question numbers are ordered by difficultly, so we might expect that students would start with low-numbered (easy) questions and work up to high-numbered (hard) questions. Indeed, the student (a) essentially used this pattern. In contrast, student (b) varied the question order more, including two easy-to-hard passes over the questions, while student (c) significantly varied the order in which they chose to attempt the questions.

To capture the full range of question-trajectories chosen by students, for each student we computed the Kendall rank correlation coefficient (Kendall tau)[12] between the attempt number and the question number sequences. This is a non-parametric statistic used to measure the similarity of the ordering between two sequences, so a coefficient of 1 means that a student attempted questions in a strictly increasing order of difficulty, while a coefficient of 0 means that there was no correlation between the question difficulty and the order in which it was attempted, meaning that students alternated or cycled between questions throughout the homework.

Figure 9(d) shows the histogram of Kendall tau coefficients for all students, with the three representative students from Figures 9(a)–(c) marked. This shows that there was a wide range of behaviors, with some students moving fairly linearly through the questions (tau coefficients near 1) and others moving in a less-linear order (tau coefficients towards 0). The average student had a tau coefficient of 0.65, showing that representative student (b) is probably a reasonable model of how a typical student chose to order their question attempts. This indicates that PrairieLearn is at least somewhat successful at inducing students to vary the order in which they solve the homework questions, rather than simply working through the homework in linear order, and it suggests that PrairieLearn may be encouraging spaced-repetition learning.

5. Conclusions

The PrairieLearn adaptive online homework system was developed and implemented within a large introductory mechanics class (*Introductory Dynamics*) at the University of Illinois at Urbana-Champaign in Spring 2014, and used again in subsequent semesters, as well as serving as the platform for a new Computerized Testing facility[14]. Student feedback was collected in Spring 2014 ($N = 408$ students) and showed improved student ratings for the online homework system, especially regarding its utility for exam preparation.

Student interaction data was collected in Fall 2014 ($N = 194$ students) and analyzed to understand the usage patterns of PrairieLearn by students. We observed a high degree of repetition on questions, which suggests that students were practicing to mastery on specific skills, and there was also a significant degree of out-of-order question attempts, which may indicate that students were gaining some benefits from spaced repetition.

We found that students choose highly individualized patterns to work through the homework questions, and the number of questions attempted and fraction of correct attempts varied significantly among students. Student exam success is not correlated with the number of question attempts that students make, but it is correlated with the fraction of attempts that are correct (correlation coefficient of 0.51).

Of particular note is the fact that students did not persist with questions that they could not solve within the first few (three or fewer) attempts. They either successfully solved the question within the first few attempts, and were then prepared to solve it again repeatedly, or they gave up on that question. This suggests that additional scaffolding may be necessary in such cases, and that

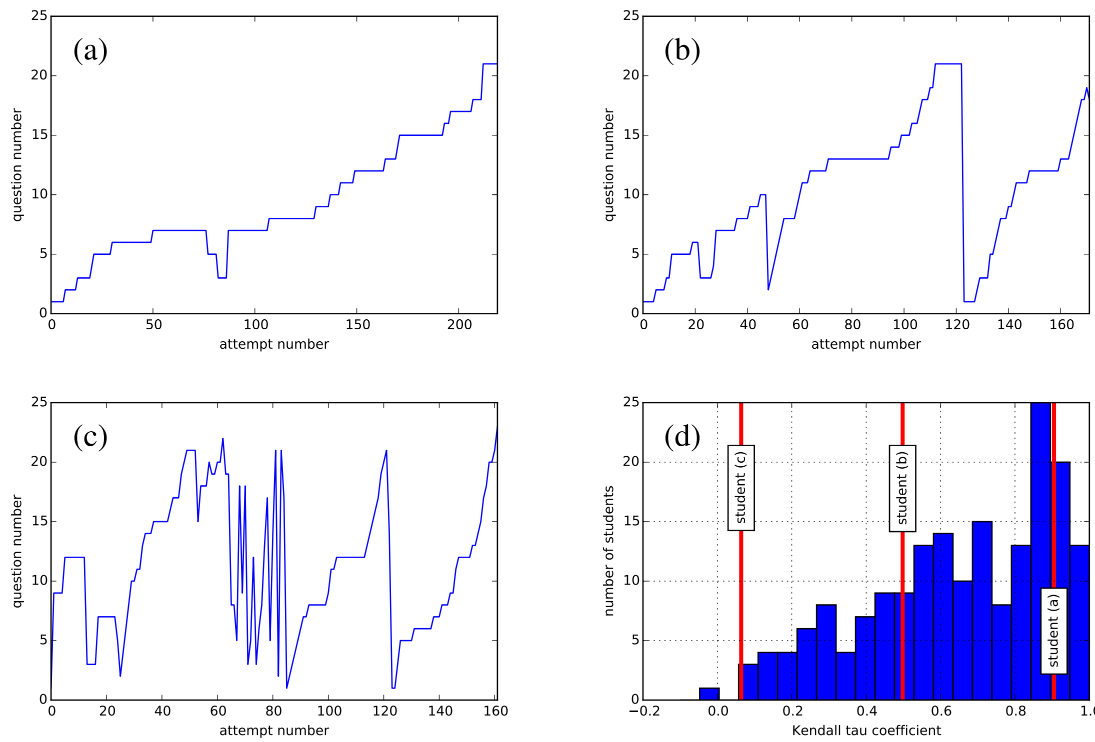

Figure 9: Question number trajectories on Homework 4 for three different students, with Kendall tau coefficients of (a) 0.91, (b) 0.50, and (c) 0.06. (d) Distribution of Kendall tau coefficients for all students on Homework 4 (mean 0.65, standard deviation 0.25), with vertical lines showing the coefficients of the three example students.

PrairieLearn should identify students who are about to give up on a question and provide them with additional assistance or incentive to continue attempting that question.

Acknowledgments. This work was supported by the College of Engineering and the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign as part of the Strategic Instructional Initiatives Program (SIIP), as well as by the National Science Foundation (NSF) award DUE-1347722.

References

[1] J. R. Anderson. *Learning and memory: An integrated approach*. John Wiley and Sons, second edition, 2000.

[2] A. D. Baddeley. *Human Memory: Theory and Practice*. Psychology Press, second edition, 1997.

[3] F. B. Baker and S.-H. Kim. *Item Response Theory: Parameter estimation techniques*. Marcel Dekker, second edition, 2004.

[4] L. Crowley and G. L. Herman. Using faculty communities to drive sustainable reform: Learning from the Strategic Instructional Initiatives Program. In *ASEE 2014: Proceedings of the American Society for Engineering Education 121st Annual Conference and Exposition*, 2014. Paper ID #9052.

[5] J. L. Davis and T. McDonald. Online homework: Does it help or hurt in the long run? In *ASEE 2014: Proceedings of the American Society for Engineering Education 121st Annual Conference and Exposition*, 2014. Paper ID #8813.

[6] R. L. Greene. Repetition and spacing effects. In H. L. Roediger, III, editor, *Cognitive Psychology of Memory*, volume 2 of *Learning and memory: A comprehensive reference*, pages 65–78. Elsevier, 2008.

[7] T. R. Guskey. Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery". *Journal of Advanced Academics*, 19:8–31, 2007.

[8] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition, 2011.

[9] J. M. Hilbe. *Logistic Regression Models*. Chapman & Hall/CRC Press, 2009.

[10] K. A. Lack. Current status of research on online learning in postsecondary education. Ithaka S+R, 2013.

[11] F. M. Lord. *Applications of item response theory to practical testing problems*. Erlbaum, 1980.

[12] A. V. Prokhorov. Kendall coefficient of rank correlation. In M. Hazewinkel, editor, *Encyclopedia of Mathematics*. Springer, 2001.

[13] M. West and G. L. Herman. Sustainable reform of "Introductory Dynamics" driven by a community of practice. In *ASEE 2014: Proceedings of the American Society for Engineering Education 121st Annual Conference and Exposition*, 2014. Paper ID #10519.

[14] C. Zilles, R. T. Deloatch, J. Bailey, B. B. Khattar, W. Fagen, C. Heeren, D. Mussulman, and M. West. Computerized testing: A vision and initial experiences. In *ASEE 2015: Proceedings of the American Society for Engineering Education 122nd Annual Conference and Exposition*, 2015.