ATLANTA
120th ASEE Annual Conference & Exposition
FRANKLY, WE DO GIVE A D*MN
June 23-26, 2013

Paper ID #7074

Using Educational Data Mining to Identify Correlations Between Homework Effort and Performance

Mr. James Herold, University of California, Riverside

James Herold earned his B.S. in computer science at California Polytechnic State University, Pomona in 2004. He is currently a Ph.D. student in computer science at the University of California, Riverside where he is researching applications of Data Mining in Educational Research Methods.

Dr. Thomas Stahovich, University of California, Riverside

Dr. Stahovich received his B.S in Mechanical Engineering from UC Berkeley in 1988. He received his S.M. and Ph.D. from MIT in 1990 and 1995 respectively. He conducted his doctoral research at the MIT Articial Intelligence Lab. After serving as an Assistant and Associate Professor of Mechanical Engineering at Carnegie Mellon University in Pittsburgh, PA, Dr. Stahovich joined the Mechanical Engineering Department at UC Riverside in 2003 where he is currently a Professor and Chair. His research interests include pen-based computing, educational technology, design automation, and design rationale management.

Mr. Kevin Rawson, University of California, Riverside

Mr. Rawson received his B.S.E. in Mechanical Engineering and B.S. in Mathematics from Walla Walla University in 2001. He received his M.S. in Mechanical Engineering from UC Riverside in 2005, where he currently is working towards his Ph.D.

Research interests include sketch understanding, machine learning, pen-based computing, and educational informatics.

Using Educational Data Mining to Identify Correlations Between Homework Effort and Performance

Abstract

Homework has long been a cornerstone of education, but is it actually worthwhile for a student to put effort into homework? In this paper we present novel techniques for examining correlations between students' effort on homework and their performance in a course. Students enrolled in a Mechanical Engineering Statics course at the University of California, Riverside were given Livescribe™ digital pens with which they completed their coursework, producing an electronic, time-stamped record of all of their work. We computed numerical features from these records to estimate the effort students expended on each homework assignment. We used these features to predict student performance on a number of measures, such as homework, quiz, and exam scores, and show that these effort-based features can explain up to 39.9% of the variance in student performance (i.e., $R^2 = 0.399$). These effort-performance correlations offer insight into the types of transfer that occurs from homework to exam problems. Additionally, these results serve as a measure of the effectiveness of homework problems, providing instructors with a principled method for improving homework assignments for future course offerings.

Introduction

Homework serves a number of purposes. It provides students with an opportunity to practice methods they have learned in the classroom, familiarizes them with new material before it is covered in lecture, and helps them synthesize concepts and apply them in new ways. Despite its widespread use, there is contention as to whether homework leads to better course performance. Numerous studies have examined the existence of correlations between a student's effort on homework and performance in a course, yet the results of these studies are mixed.

Variations in the nature of these studies may partially account for these inconsistencies. For example, these studies vary in the grade-level of the students, the type of homework assigned, and the subject matter. Additionally, bias and inconsistencies in the measurement of homework effort may also confound the results. Most previous work relies on the students themselves, or their parents, to report the amount of time spent on homework.

In our work, by contrast, we rely on more precise and objective measures of homework effort. Our analysis is enabled by the unique way in which we capture students' ordinary problem-solving processes. In the winter quarter of 2012, students enrolled in a Mechanical Engineering Statics course at the University of California, Riverside were given Livescribe™ digital pens with which they completed their coursework, producing a digital record of all their work. We compute numerical features from these records which estimate the effort students expended on each homework assignment. We use these features to predict students' performance on a number of measures, including homework, quiz, and exam scores. We show that these effort-based features explain up to 39.9% of the variance in the students' performance. These results have several pedagogical implications. The correlations we identify provide insights into the types of transfer students make from homework to exam problems. The results can also be used to evaluate the effectiveness of homework assignments, allowing instructors to improve homework assignments for future course offerings. Lastly, our analyses can be used to identify patterns of homework effort exhibited by students who perform well in the course.

Related Work

Mayer[10] discusses two types of tests of student learning: retention and transfer tests. The former requires a student to solve a problem that is similar to one he or she has already studied, while the later requires the student to apply existing knowledge to a novel problem. Mayer notes that students typically perform better on routine problems (retention) than on non-routine ones (transfer). This distinction between retention and transfer provides the context for the results we present below.

There are numerous studies which seek to identify correlations between performance and the amount of effort spent on homework[2,6,3,11,13]. However, the results of these studies are inconsistent. Some indicate that a significant correlation exists between homework effort and performance, while others find no such correlation, or even negative correlations. However, the results of these studies do suggest that subject matter and grade-level moderate the effort-performance correlations. Typically, correlations are found to be stronger in some subjects than in others. Also, the correlation is typically larger for higher grade levels. Cooper et al.[4] compared the results of each of these studies and found an average correlation of $r = 0.14$, with a range from -0.25 to 0.65. The correlations found in our paper are typically stronger than those found in this prior work.

Cooper et al.[4] summarize this inconsistency in findings: "To date, the role of research in forming homework policies and practices has been minimal. This is because the influences on homework are complex, and no simple, general finding applicable to all students is possible." The complexity of the influences of homework should by no means prohibit research into its effectiveness though. On the contrary, this complexity underlies the importance of developing general, data-driven techniques that can easily be applied in any course to determine the effectiveness of a particular homework assignment.

In the present research, we apply Educational Data Mining techniques to identify correlations in students' work in a data-driven way. Educational Data Mining is a nascent research field in which Data Mining techniques are applied to educational data in order to make discoveries about how students learn. Recent work in this field typically focuses on data extracted from either Learning Content Management Systems (LCMS) or intelligent tutoring systems (ITS). For example, Kinnebrew and Biswas[9] examined students' interactions with the Betty's Brain ITS. This system logged the actions students took while working with it. Sequence mining techniques were then applied to identify series of actions that correspond to productive and unproductive learning behaviors. Romero et al.[12] applied Data Mining techniques to data collected with the Moodle LCMS which records detailed logs of students' interactions, such as viewing and submitting assignments. These logs were mined for rare association rules describing patterns that appear infrequently in the data. The resulting rules were then manually inspected to identify fringe behaviors exhibited by students.

Oviatt et al.[14] examined computer interfaces for completing geometry problems and found that "as the interfaces departed more from familiar work practice..., students would experience greater cognitive load such that performance would deteriorate in speed, attentional focus, meta-cognitive control, correctness of problem solutions, and memory." Thus it is important for Educational Data Mining to be applied to data collected under ordinary work conditions.

Van Arsdale and Stahovich[15] have demonstrated a correlation between the temporal and spatial organization of a student's handwritten exam solutions and the correctness of the work. The work was captured with digital pens. The organization was characterized by a set of quantitative features computed from the digitally recorded ink. On average these features accounted for 40.0% of the variance in students' performance. Similarly, Herold and Stahovich[8] applied Data Mining techniques to identify how self-explanation affected students' problem solving processes. The study included an experimental group which provided handwritten self-explanation for the major steps in each homework solution, and a control group which did not. Digital copies of the students' handwritten homework were mined for commonly occurring patterns, revealing that students who generated self-explanation solved problems more like an expert than those who did not. In this work, we employ similar, data-driven techniques to identify correlations between students' homework effort and performance.

Data Collection

In the winter quarter of 2012, over 120 students enrolled in an undergraduate Mechanical Engineering course on Statics at UC Riverside were given Livescribe™ digital pens which they used to complete their coursework. In this way, we collected a digital, time-stamped record of six homework assignments, seven quizzes, two midterms, and the final exam. Most homework assignments comprised eight problems, each of which would take approximately 30 minutes to solve. An example of a typical problem is provided in Figure 1. Assignments were typically due one week after they were assigned. Our present analysis excludes data from the first two homework assignment and quizzes as they concerned basic math skills, rather than equilibrium analysis, which is the primary focus of the course.

Computing an Estimation of Student Effort

Here we describe the two types of novel quantitative features we use to estimate students' effort on homework. The *overall-effort* features are coarse-grained, and characterize the total effort a student spent on a particular assignment. The *per-problem* features are fine-grained, and characterize the amount of effort spent on each individual problem.

Overall-Effort Features and Performance Models

The overall-effort features characterize the distribution of effort a student expends on his or her homework assignment. For example, some students may begin an assignment early and put substantial work into it each day, resulting in several homework "episodes." Conversely, other students may put off the homework until shortly before it is due, resulting in a single, large homework episode.

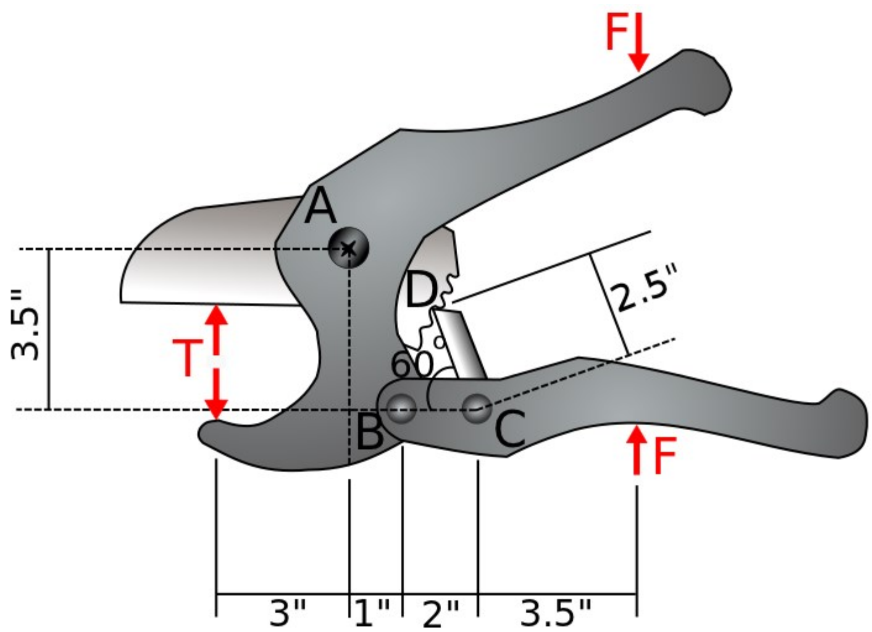

Figure 1: *Typical homework problem from the Statics course. The problem statement reads: "The device shown is used for cutting PVC pipe. If a force, F = 15 lb, is applied to each handle as shown, determine the cutting force T. Also, determine the magnitude and the direction of the force that the pivot at A applies to the blade."*

To compute the overall-effort features, we first create a time-series representing the effort a student exerted on an assignment . The series begins with the first pen stroke written and ends with the last. This time span is divided into five-minute intervals. Each interval is characterized by the amount of ink written, which is defined as the distance the pen tip travels on the paper during that interval. In this way, effort is characterized by the amount of writing rather than simply the amount of time elapsed.

Figure 2 shows a typical effort time-series. Effort time-series are typically flat and punctuated with a few, large episodes of activity. We compute four features from each time-series. The first feature is the total amount of ink written, which characterizes the total effort spent on that assignment. The remaining three features characterize the distribution of this effort. To compute

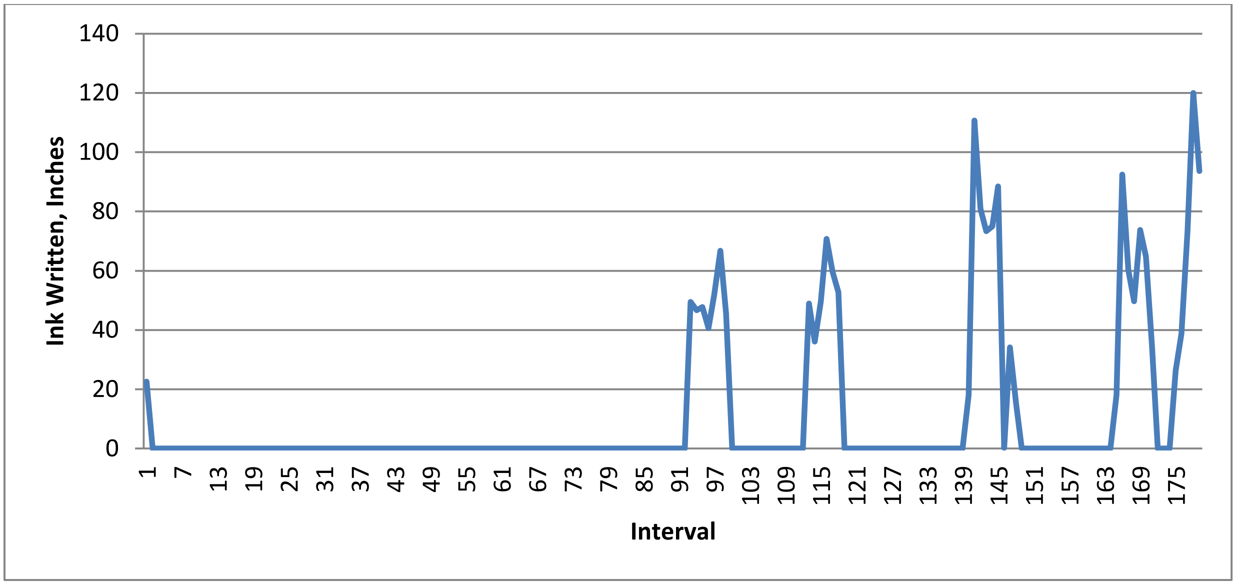

Figure 2: *Typical effort time-series of a single student on a single assignment. The abscissa denotes the index of the five-minute intervals, not the actual time stamp of the interval. For example, interval one comprises the first five minutes of the students problem solution. The ordinate denotes the total ink written during a given five-minute interval.*

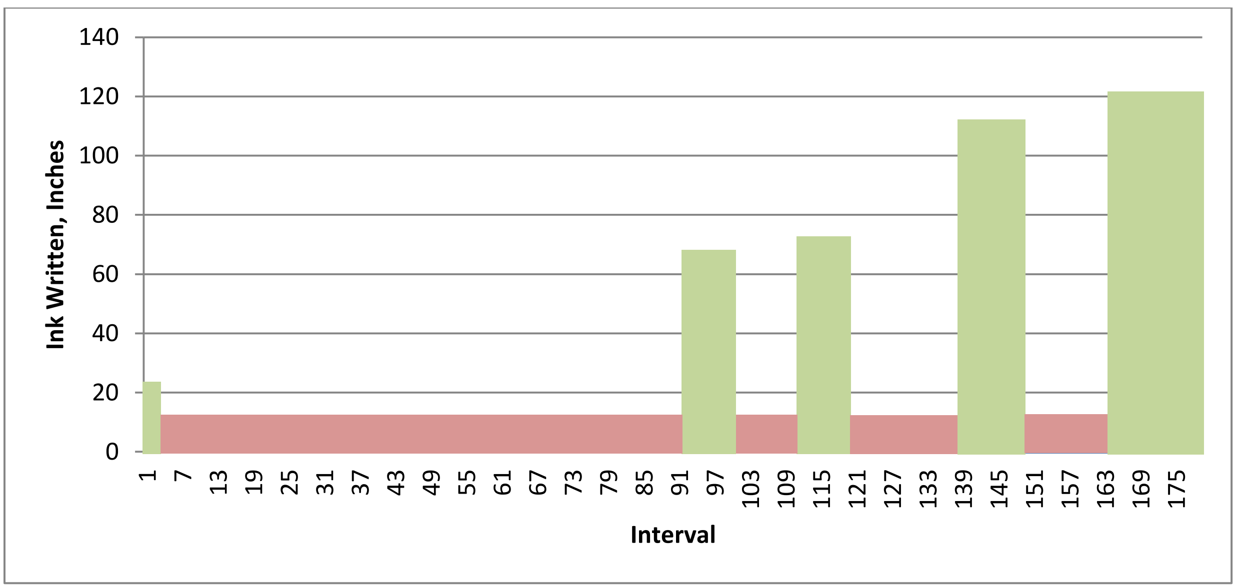

Figure 3: *Effort time-series previously shown in Figure 2 with active (green regions) and inactive (red regions) episodes identified.*

these features, we first identify *active* episodes, in which a student is writing, and *inactive* episodes in which no writing occurs. Each contiguous sequence of non-zero intervals (i.e,. intervals containing writing) forms an active episode. To prevent small breaks in writing from splitting an episode, active episodes may contain subsequences of up to two zero-valued intervals. Thus a break of ten minutes or less does not break a problem-solving episode. All remaining contiguous sequences of zero-valued intervals are identified as inactive episodes. The effort time-series is characterized by the number of active episodes, the average length of the active episodes, and the average length of the inactive episodes. Figure 3 shows the active and inactive episodes for the effort time-series from Figure 2.

We used the four overall-effort features to construct models relating students' effort on a particular assignment to performance on that assignment. We computed these models using the linear regression package in the WEKA Data Mining Software suite[7]. Figure 4 presents the coefficient of determination for the models constructed for each homework assignment. WEKA's linear regression package employs a greedy feature selection algorithm. Features are removed from the model until there is no improvement in the error estimation, as determined by the Akaike information criterion[1]. The features selected for each homework model are presented in Table 1.

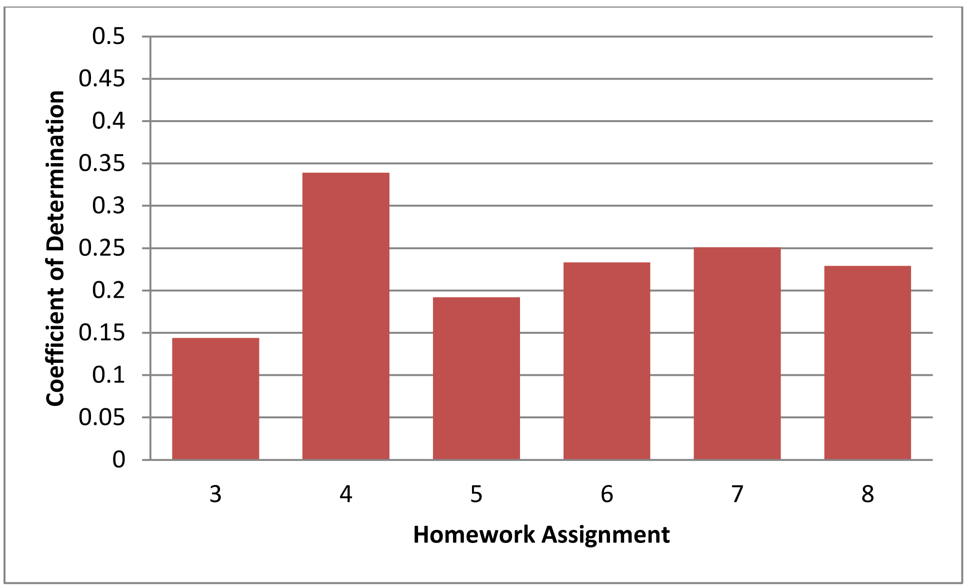

Figure 4: *Coefficient of determination for linear regression models relating overall-effort features to homework performance. The average coefficient of determination is 0.231.*

	Avg. Active Length	Avg. Inactive Length	No. Active Intervals	Ink Written, Inches
Homework 3			Selected	
Homework 4			Selected	
Homework 5		Selected		Selected
Homework 6	Selected			Selected
Homework 7		Selected		Selected
Homework 8			Selected	

Table 1: *Overall-effort features selected for linear regression models for homework performance.*

We also used WEKA's implementation of the expectation maximization (EM) clustering[5] algorithm to group students by similarities in both the effort they exerted on a homework assignment and they performance they achieved on it. The clusters identified for each assignment are listed in Table 2 and 3. Each cluster is characterized by the average and standard deviation of the homework grade (the maximum grade is 10.0) and the overall-effort features of the data points contained in that cluster.

Per-problem Features

We use the overall-effort features to examine the relationship between the total effort on an assignment and performance on that assignment. Here we examine how effort on individual homework problems relates to performance on subsequent homework assignments, exams (midterms and final), and quizzes. We estimate the effort on a single homework problem as the total time during which the pen is in contact with the paper. The time between pen strokes is not included in this value. We once again employ the linear regression package available in WEKA to compute regression models. The resulting coefficients of determination are shown in Figure 5. In the figure, the coursework is listed in the order it was completed. For example, quiz six was completed just after homework seven was due. Furthermore, a model using the effort on each of the problems from homework assignments three through seven predicts performance on quiz six with a coefficient of determination of 0.32.

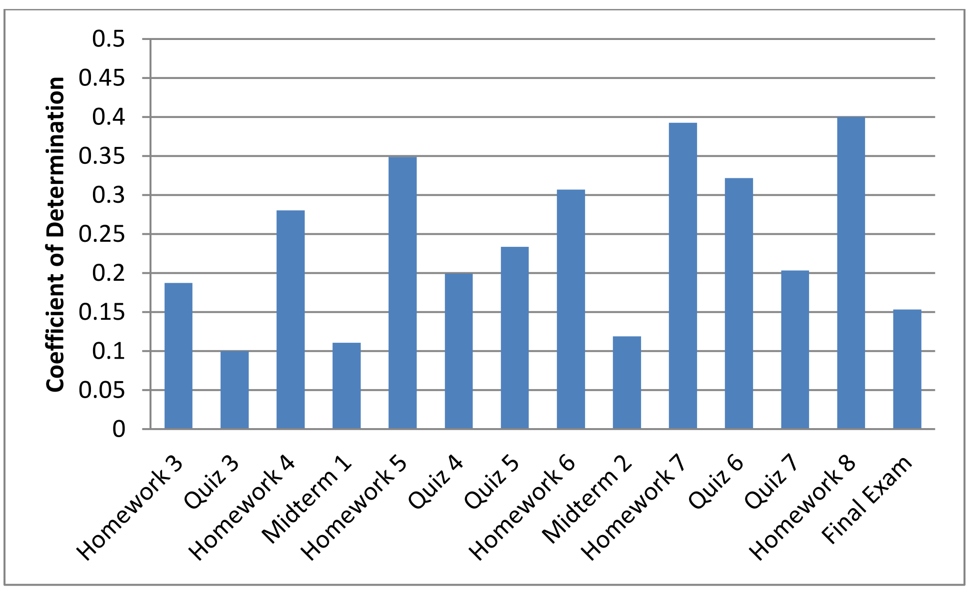

Figure 5: *Coefficients of determination of models using per-problem features to predict performance on coursework. The average coefficient of determination is 0.239.The coursework is listed in the order completed.*

In a similar way, we used the per-problem effort features to predict performance on individual problems on the midterm and the final exams. The results are shown in Figure 6. The features selected by WEKA's greedy feature selection algorithm provide insights about learning transfer. For example, the features selected for predicting performance on the first problem of the first midterm were effort on homework assignment three, problem four; effort on homework three, problem five; and effort on homework four, problem five. (Because of the large number of features we consider, space constraints prevent inclusion of the complete feature selection results corresponding to Figures 5 and 6.)

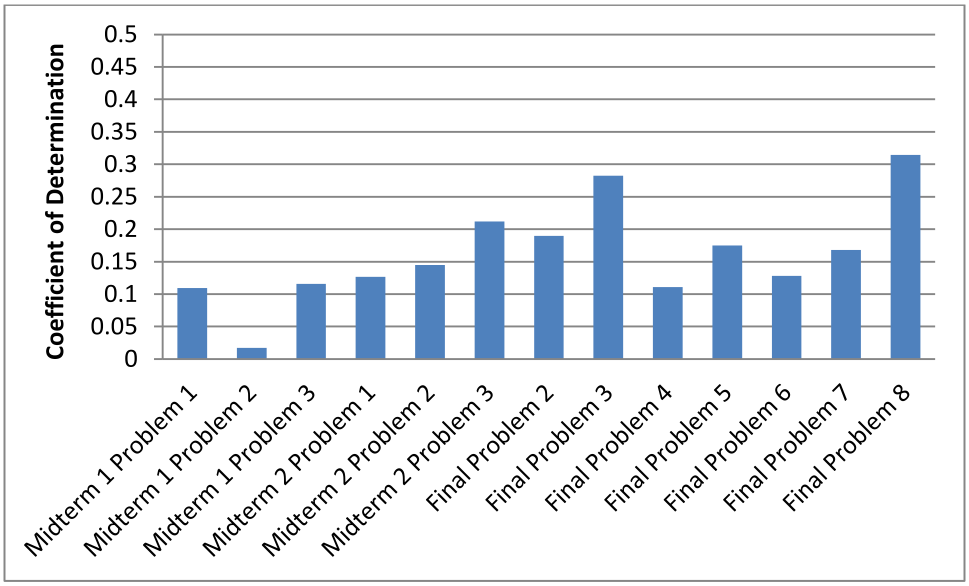

Figure 6: *Coefficients of determination of models using per-problem features to predict performance on individual exam problems. The average coefficient of determination is 0.161. (The first problem of the final exam concerned professional ethics question and thus was excluded from the analysis.)*

Discussion

It is important to note that our effort features capture only a portion of the effort expended by students on studying. Other elements of studying, such as the amount of time spent reading the textbook or working on scratch paper, are not captured by the digital pens we use. However, we believe that the amount of time spent problem solving on homework provides a useful measure of a student's effort in a course.

The results of the linear regression analysis of the overall-effort features indicate that students' effort does account for a considerable portion of the variance in performance. For example, in the best case, the effort-based features accounted for 33.9% of the variance in performance on homework. This correlation is considerably stronger than that found in previous studies. Interestingly, the feature selection results indicate that the number of active intervals and the total ink written are the most important features. Each is selected for the models for at least half of the homework assignments. This suggests that the more often a student sits down to work on an assignment and the more writing he or she does, the more likely it is that the student will do well on that assignment.

Each of the linear regression models showed a positive correlation between performance and each effort feature, indicating that the more time a student spent on a particular problem the better he or she performed. This may demonstrate that students who spent more time on their homework were better prepared for the exam problems. It is important to note that this may not always be the case. It is entirely possible that a particular effort feature could negatively correlate with performance. Such a case may indicate that a student spent a large amount of time on a problem as because he or she had difficulty understanding it and as a result that student did not perform well on related exam problems.

The clustering results reveal a similar story to the linear regression analysis. The clusters with the highest average grade are typically those which also have the highest average number of active intervals. The clustering results serve as an easy-to-read summary of typical solution behaviors exhibited by students on a particular assignment. Instructors can use these sorts of results to quickly determine which groups of students are spending a sufficient amount of time on the homework. More importantly, this analysis reveals just how much time is needed to do well on an assignment. This will enable an instructor, for example, to identify when a large number of students are performing poorly on a problem despite spending a great deal of effort on it, a strong indication that a widespread misconception or difficulty exists in the class.

The results of the per-problem linear regression analysis reveal that the amount of effort spent by students on individual homework problems can account for up to 39.9% of the variance of students' performance on subsequent homework assignments. Furthermore, when the per-problem features are used to predict performance on individual exam problems, they can account for up to 31.4% of the variance in that grade. This is an interesting result as these features do not consider the semantic content of the students' solutions.

Cluster		Homework 3			Homework 4			Homework 5			
		1	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*4*
Grade	Mean	5.35	6.2	7.23	3.53	3.75	6.87	5.04	5.56	2.45	1.61
	Std.	2.92	3.3	3.08	3.57	3.36	1.48	1.7	2	3.11	2.16
Avg. Active Length	Mean	393.86	14.8	100.16	654.97	56.78	14.86	109.08	12.23	873.58	23.12
	Std.	412.1	9.83	45.42	115.82	64.93	9.97	84.55	8.3	303.81	36.64
Avg. Inactive Length	Mean	779.13	80.72	76.69	636.38	83.58	174.18	93.79	164.16	855.98	276.44
	Std.	273	82.56	41.81	119.52	86.68	115.59	71.28	118.78	308.34	217.46
No. Active Intervals	Mean	2.5	2.88	2.03	2	2.06	4.5	1.96	4.81	2	2.93
	Std.	0.5	1.77	0.18	1.81	0.95	1.61	0.62	1.69	1.77	1.03
Ink (Inches)	Mean	905.9	1045.4	1216.0	804.8	1001.1	166.9	1088.9	176.5	675.4	925.1
	Std.	247.4	692.7	1168.4	358.6	652.1	758.5	680.2	756.6	401.6	539.6

Table 2: *The average and standard deviation for each of the EM clusters for homework assignments three to five.*

Cluster		Homework 6				Homework 7		Homework 8			
		1	*2*	*3*	*4*	*1*	*2*	*1*	*2*	*3*	*4*
Grade	Mean	5.57	8.02	3.89	8.23	6.38	7.36	7.61	8.09	0.78	8.81
	Std.	2.97	1.68	4.04	1.54	3.47	2.94	1.37	2.37	0.41	1.78
Avg. Active Length	Mean	973.42	10.93	97.08	33.38	10.02	686.56	650.91	24.07	73.21	26.36
	Std.	408.61	5.16	86.93	22.12	6.26	980.46	343.74	32.02	140.13	31.26
Avg. Inactive Length	Mean	1149.78	168.8	109.29	27.02	357.31	658.08	628.91	18.98	100.74	183.19
	Std.	152.99	139.52	87.22	17.19	416.62	983.1	340.41	24.58	165.08	159.16
No. Active Intervals	Mean	2.14	4.74	2.07	1.75	3.68	2	2	1.24	1.43	3.01
	Std.	0.34	1.61	0.79	0.76	2.3	2.17	1.07	0.43	0.64	1.05
Ink (Inches)	Mean	1188.2	660.3	928.1	1290.1	1237.1	1246.3	769.6	610.0	486.5	1119.0
	Std.	297.1	686.4	602.6	6124.7	667.1	555.9	343.8	202.8	293.3	342.8

Table 3: *The average and standard deviation for each of the EM clusters for homework assignments six to eight.*

The per-problem features selected by the final model of each linear regression are an indication of the importance of individual homework problems. (In the present data, all of these features were positively correlated with performance.) The selected features reveal which homework problems lead to success in learning particular concepts in the course. More specifically, they reveal the transfer taking place from particular homework problems to particular exam problems. Consider, for example, the first problem of the first midterm, shown in Figure 7. The amount of effort exerted on homework three, problem four was one of the three features selected to predict the performance on this midterm problem. Interesting, the midterm problem can be considered a rotated version of the homework problem. This clearly shows students transferring knowledge from the homework problem to solve the midterm problem.

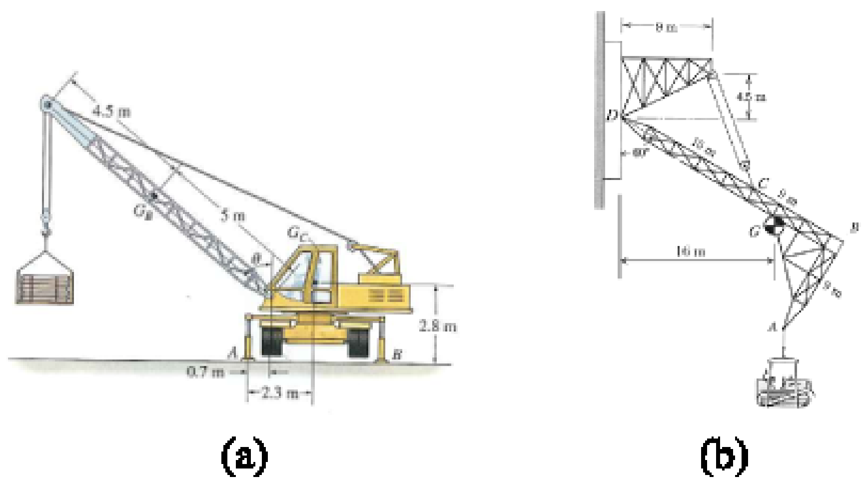

Figure 7: *Homework three, problem four (a) and midterm one, problem one (b).*

These results indicate practical changes instructors can make to homework assignments. Namely, this suggests that exam problems which comprise simple extensions to homework problems can be used to identify students' ability to transfer knowledge. Instructors should examine students' performance on the homework problems that will be similar to upcoming exam problems. If students are spending insufficient effort on these problems, they should be further examined in class.

This analysis can serve as an invaluable tool to the instructor of a course. Using it, the instructor may review which features (i.e., homework problems) are selected in the per-problem models, identify the types of transfer students made, and use that knowledge to shape exam and homework problems in future course offerings. A manual analysis of the transfer revealed by our present data will be an important element of future work on this project.

Overall, the linear regression analysis results for both the overall-effort and per-problem models provide correlations that are much stronger than those found in prior work. As mentioned earlier, those studies typically relied on either the students or their parents to report the amount of time spent working on each homework assignment. The Livescribe™ digital pens provide a more reliable measure of the amount of time students spend on their homework assignments which may account for the higher coefficients of determination we obtain.

Conclusion

In this paper, we have presented novel, data-driven methods for assessing students' homework habits in a Mechanical Engineering course. These methods are enabled by our unique data set. In the winter quarter of 2012, over 120 students in a Mechanical Engineering course completed all coursework with Livescribe™ digital pens, producing a time-stamped, electronic record of their work. Using this record, we computed a number of features which characterized the effort that students exerted when solving their homework assignments.

There were two major types of features computed: *overall-effort* features and *per-problem* features. Four overall-effort features were computed which characterized both the amount and distribution of effort exerted on a single assignment: the total amount of ink written in an assignment; the number of active problem-solving episodes; the average length of the active problem-solving episodes; and the average length of the inactive episodes. These features were used to predict the performance on the homework assignment from which they were computed. These overall-effort features explain up to 33.9% of the variance in students' performance on a particular assignment. Additionally, these features and the homework assignment grade were used as input to the EM clustering algorithm. This algorithm identified groups of students who both displayed similar effort behaviors and assignment performance. These groupings may be used as behavior-performance profiles that are valuable feedback for an instructor.

The per-problem features comprised the amount of time spent writing the solution to each problem of a homework assignment. These features were used to predict performance on subsequent homework assignments, quizzes, and exams. These features accounted for up to 39.9% of the variance on a particular item. Additionally, the per-problem features were used to predict performance on individual exam problems. They accounted for up to 31.4% of the variance of the performance on individual problems. More importantly, the features selected by these linear regression models provide important insights for the instructor, indicating which of the homework problems lead to good performance on the exam questions. By analyzing which homework problems most account for the variance on a particular exam problem, instructors may identify the types of transfer students make from homework to exam problems. This information is an invaluable source of feedback for the instructor as well as a guide to how homework and exam problems should be designed for future course offerings.

Acknowledgements
The authors gratefully acknowledge support for this work provided by the National Science Foundation via award No. 0935239.

References

[1] Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 1974.

[2] Antonek, J. L. Interactive homework: Linking Parents to the Foreign Language Learning of Their Children. *Dissertation Abstracts International*, 57, 5084, 1996.

[3] Bowen, N. K., and Bowen, G. L. The mediating role of educational meaning in the relationship between home academic culture and academic performance. *Family Relations*, 47, 45–51, 1998.

[4] Cooper, H., Robinson, J., and Erika, P. Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003. *Review of Educational Research*, 76, 1-62, 2006.

[5] Dempster, A., Laird, N., and Rubin D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39, 1-38, 1977.

[6] Deslandes, R., Royer, E., Potvin, P., and Leclerc, D. Patterns of home and school partnership for general and special education students at the secondary level. *Exceptional Children*, 65, 496–506, 1999.

[7] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11, 2009.

[8] Herold, J. and Stahovich, T. Characterizing Students' Handwritten Self-Explanations. In *Proceedings of American Society for Engineering Education Annual Conference & Exposition*, 2012.

[9] Kinnebrew, J. and Biswas, G. Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution. In *Proceedings of the Fifth International Conference on Educational Data Mining*, 2012.

[10] Mayer, R. Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 1998.

[11] Pezdek, K., Berry, T., and Renno, P. A. Children's mathematics achievement: The role of parents' perceptions and their involvement in homework. *Journal of Educational Psychology*, 94, 771–777, 2002.

[12] Romero, C., Romero, J., Luna, J., and Ventura S. Mining Rare Association Rules from e-Learning Data. In *Proceedings of the Third International Conference on Educational Data Mining*, 2010.

[13] Singh, K., Granville, M., and Dika, S. Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research*, 95, 323–332, 2002.

[14] Oviatt, S., Arthur, A., and Cohen, J. Quiet interfaces that help students think. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, 191–200, 2006.

[15] Van Arsdale, T. and Stahovich, T. Does Neatness Count? What the Organization of Student Work Says About Understanding. In *Proceedings of American Society for Engineering Education Annual Conference & Exposition*, 2012.