# AC 2007-1727: DATA-MINING AN ONLINE HOMEWORK SYSTEM

**Andrew Bennett, Kansas State University**

**Eric Lawrence, Kansas State University**

**Genevra Neumann, Northern Iowa University**

**Elena Verbych, Kansas State University**

**Steve Warren, Kansas State University**

# Data-Mining an Online Homework System


## Abstract

Online homework systems are becoming increasingly popular since (when they work) they are convenient for both faculty and students. Systems that rely on mechanical grading are naturally best adapted to more mechanical types of problems, raising issues of whether an increasing reliance on such systems will privilege the assessment of procedural knowledge over the assessment of conceptual knowledge. However, online systems naturally and efficiently capture large amounts of data about student work and data-mining techniques can be applied to evaluate conceptual understanding as well as procedural understanding, even though the prompts are all procedural. In this paper, we discuss how to use detailed analysis of procedural results captured by a locally designed online homework system (tuned for the purpose of assessing conceptual understanding) to recognize conceptual growth in classes in mathematics and the likelihood of successful transfer of this understanding to later engineering classes. Patterns that demonstrate students are wrestling with new concepts and techniques for disentangling correlations in different subjects caused by successful transfer from correlations caused by general skills are developed. While the analysis is based on our local system, the general approach and tools can be applied to other systems as long as they allow multiple attempts and retain information about unsuccessful attempts prior to the final submission.


## Introduction

Online homework is becoming a common tool in college mathematics courses, as well as other science and engineering courses. One product, WebAssign, has a list of over 300 U.S. Colleges and Universities using their system[1], and most publishers now offer online homework systems as an option with many of their texts. The popularity of online homework systems is easy to understand. For the faculty, an online homework system reduces the amount of effort spent on grading and can also reduce management issues relating to collecting, recording, and returning student papers. For the students, online homework systems allow them to work on their own schedule and receive immediate feedback on what they have done correctly and incorrectly. In addition, some systems allow students multiple attempts and extra practice compared to courses with traditional homework only. Given the practical advantages, the shift to online homework seems very likely to continue. Therefore, it is important to study how this shift may alter instruction and learning, and how teachers can best assess student learning in an online world.

Using an online homework system can influence the types of assignments that are made. Online homework problems must be of a format that can be graded by machine. This can lead to more procedurally oriented problems that have well-defined answers and for which systems can be easily generate multiple variations by simply adjusting the numbers. More conceptual problems are more likely to require open-ended solutions and are usually much more difficult to implement with such systems. Thus the shift to more online homework raises the possibility that it will be accompanied by a shift toward more

emphasis on procedural work and a decrease in emphasis on understanding the concepts being taught.

While the shift to online homework raises questions about the range of problems that will be assigned, it also provides new opportunities for understanding student learning. A natural feature of online systems is that one can track in a database how students interact with the system (though many systems store only limited information). Applying data-mining techniques to analyze this information offers the possibility of developing a better understanding of how our students are learning. In addition, student work becomes easier to track over time, allowing for longitudinal research into how success in one class correlates (or fails to correlate) with success in later classes. Of course, correlation is not causation. This is a particularly tricky issue when looking at student performance where it is likely that correlations result not from transfer of knowledge between different classes, but from the fact that a bright student is likely to do well in all his or her classes. But, since most students take service courses in mathematics specifically to prepare for later courses, understanding whether and how they transfer their learning is important.

An issue to be considered in such data-mining is the level of granularity of the results obtained. By level of granularity, we mean whether one is detecting if specific individuals are gaining conceptual understanding or whether we can just say that approximately some percentage of the class is gaining such understanding, without necessarily identifying which students were in that group. Homework assessments have traditionally been used to determine the understanding of each individual student, as have many other assessments. On the other hand, some standardized assessments, such as the NAEP[2] and TIMSS[3] data, have been designed to assess how states or even countries are doing. It is not initially obvious what level of granularity data-mining might provide, nor what level would best serve the needs of the instructor. Since data is collected for each student, one might hope for answers at the level of the individual student. However, the statistical analysis we will need to recognize conceptual growth from procedural data may require more data than we will have for individual students. This may not be a weakness of the analysis. Service courses for engineers are often taught in large lectures, and an instructor facing a class of 250 may be better served by knowing that most of the students are understanding topic A while relatively few understand topic B than by having 250 separate profiles covering each individual.

With these ideas in mind, the goals of the research reported in this paper are the following. Create an online homework system addressing procedural problems that tracks student usage carefully. Apply data-mining techniques to the data collected by the system to answer the following questions

1. Can conceptual learning be identified from analysis of student responses to procedural problems?
2. Can transfer of learning between classes be identified from analysis of online homework data?
3. At what level of granularity can these questions be answered?

**The Online System**

We developed an online system used at our school (a large Midwestern university) in Trigonometry, Calculus 2, Elementary Differential Equations, and which is currently being extended to College Algebra. These classes are taught in a lecture-recitation format with large lectures of 150-300 students. A slight variation of the system (omitting the two-stage grading) has been implemented in the Electrical Engineering department's Linear Systems class. This online system has a number of features that affect student usage. Not all these features are necessary to support our later analysis. The key issues are that students are allowed multiple different problem sets on each assignment (rather than one problem set being revised repeatedly until perfected) and that the system records unsuccessful as well as successful work. Of course, it should be stated that many people beyond the authors aided in the development of the system.

Each student gets an individual problem set. Problem statements remain the same but the coefficients will change for each student. Note that changing coefficients can sometimes change the procedures needed to solve the problem. For example, changing coefficients may change an integral from a form best handled by trigonometric substitution to one best handled by integration by parts.

Problems are procedural, but answers may be requested as numbers, formulas, or graphs. Formulas are typically evaluated at 4 points and are marked correct if they agree with the true answer at all points. In selected problems (particularly in college algebra), pattern-matching algorithms are used instead to determine if the answer is in a specified format. Graphs are produced by students using an applet which is then queried by the system to determine the key features of the student graph for grading.

Once a student logs into the system and has a problem set generated, the student keeps that problem set until he or she has submitted their answers for final grading. The student is free to print out the problems to work at some other time, returning later to input the answers. Students can also save work they have input without having it graded until later.

Grading takes place in two stages. The first time a student submits work for grading, her or she is told if each answer is correct or incorrect. If there are no errors, the paper is marked fully correct and a final grade recorded immediately. Otherwise, the student has a chance to correct any errors (the system doesn't permit students to "correct" right answers). The second time work is submitted for grading, students are told if each answer is correct or incorrect, what the correct answer is and are given a link to see how to work any problems that were incorrect (with the specific numbers of the incorrect problem). The final grade for that problem set is recorded at this time. Using two-stage grading has both pedagogical and practical advantages. Pedagogically, it encourages students to examine their work to find errors, and provides feedback during the learning process instead of just at the end. Practically, the two-stage grading gives students a chance to fix typographical errors, which increases their happiness with the system.

Students may work multiple problem sets for each assignment. Once a problem set is completed, the student may log in again and get a new problem set (with similar

problems but new coefficients). The student receives the highest grade they achieve over all problem sets on a particular assignment. Note that students don't keep correct answers for the new problem set. So if a student has 2 out of 3 problems right on the first problem set, they have to get 3 out of 3 on the second problem set to improve their grade, not just 1 out of 1.

Problem sets are usually designed to take approximately 20-30 minutes to complete, since that appears to maximize student effort and learning (long enough for the students to master the material but not so long the students get frustrated and refuse to try again if their scores are low). This typically means 3-5 problems in calculus and differential equations, with sometimes more problems in precalculus classes. Problem sets are due at midnight on specified days. Problem sets are still available after the due date and some students work problem sets for practice before each exam.

The system automatically records the students' problems, saved work, initial answers, final answers, time of all accesses of the system, ip-addresses of all accesses of the system, and whether they checked the help page for problems they missed.

**Analysis**

Since the system was first developed in Trigonometry, we have the most data in that course (a key consideration for a data-mining project). So we will focus our analysis on Trigonometry. The first question is whether it is possible to detect conceptual understanding from analysis of procedural work. Van Hiele[4] has noted that when students move from material at one conceptual level to a higher level, there is usually a sudden decrease in the speed of learning as the students struggle with the new material. Some students then speed up as they master the higher level, while others continue to struggle. We adopt this approach to searching for evidence that students are encountering and mastering (or not) new concepts.

Measuring speed is a little more complicated than it may seem. While the system records when students receive problems and when they submit their answers, there is no way to tell how much time students spent actually working on problems between these two events. In fact, frequently it is clear that students were not working for significant lengths of time between getting the problems and submitting their answers. A student who gets the problems one day and submits the answers two days later may reasonably be assumed to have printed out the problems and worked on them at some point, but not for 48 hours straight. On the other hand, some students will work several problem sets in quick succession, in which case we can reasonably assume most if not all the time from getting the problems to submission of answers is spent in working the problems. Since one of the design criteria is that problem sets should take roughly the same amount of time to complete, we use samples where it is reasonable to measure the length of time working problems to calibrate how long the problem sets are and then use the number of problem sets attempted, which can be well-measured for all students, to calculate speed. Hence we take the inverse of the number of problem sets attempted on average on each assignment

as the speed with which students work an assignment. Speeds for the 12 assignments over the course of the semester are shown in the graph below.

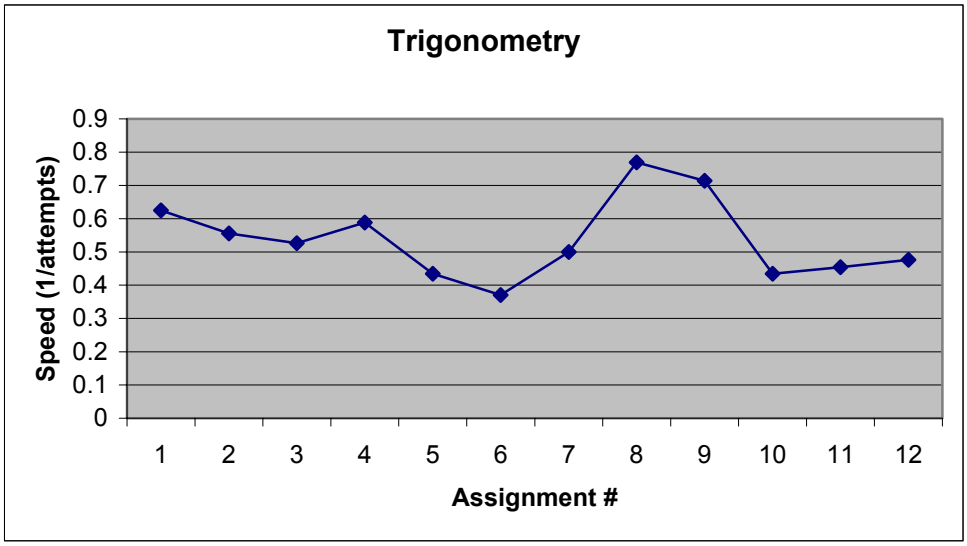


Figure 1

A short aside on statistical significance of our results is appropriate at this point. The data above is drawn from the Fall '06 semester, with the number of students completing each assignment ranging from 238 to112. There was significant falloff in the number of students completing the last assignment – we have found that first and last assignments in many classes do not provide consistent data as student performance is strongly affected by other factors at the beginning and end of the semester. While not every shift in the curve above is statistically significant (for example, the changes in speed between assignments 2-4 are insignificant), the changes we will discuss in this article are statistically significant. Perhaps more important from the standpoint of establishing that the points being discussed are real and not just the result of random variation, the features discussed in this article have been observed consistently over several years. The graph of speeds for other semesters looks very similar to the graph above.

In the graph above, the first 4 assignments deal with right angle trigonometry and introduce the unit circle. Assignments 5 and 6 deal with symmetry properties of trig functions and require the students to work with function properties of sine and cosine, not just treating these as properties of a triangle. Assignment 7 introduces complex numbers and assignment 8 applies trig identities to the solution of triangles. Assignments 9-12 cover analytic geometry starting with a review of lines in assignment 9 and then three assignments on graphing conic sections in 10-12, where the primary technique is completing the square.

Looking at the data, we see the speed starts out relatively constant, dips at assignment 5 before recovering later, then falls more precipitously at assignment 10. Since assignment 5 moves from thinking of trig functions as properties of angles to thinking of them as primary objects themselves with their own symmetry property, it appears the students

may be struggling with the concept of function at this point. Assignment 10 introduces new material for analytic geometry, hence the slowdown here may also be related to difficulties with new concepts. Of course, these dips have other possible explanations as students might reasonably allot more or less time for assignments as the semester progresses. We have found a better signal of conceptual difficulty is by comparing the percentage of students who get A's on each assignment (defined as 90% or better) with the mean score on each assignment. This data is graphed below.

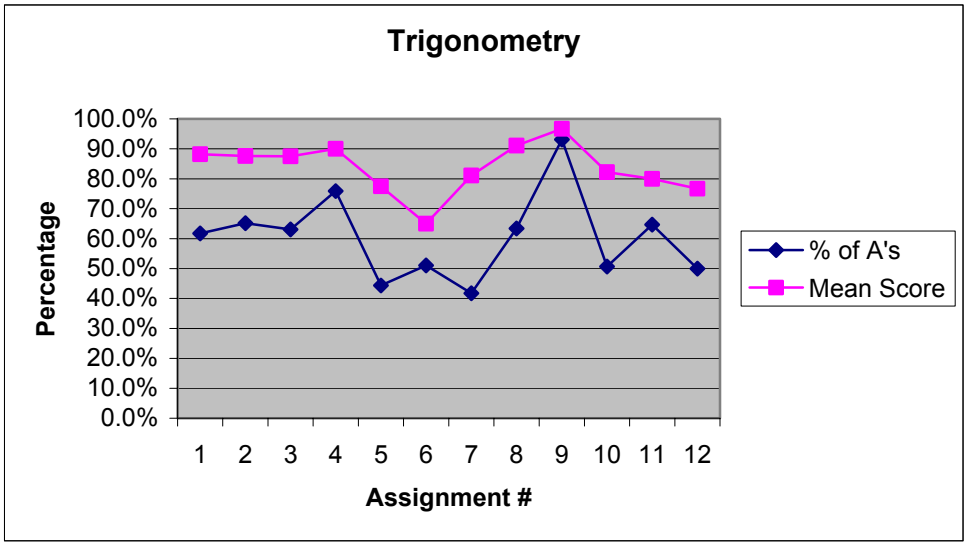


Figure 2

Not surprisingly, in general the percentage of A's tracks the mean score. However, there are a couple of places the two move in different directions. In particular, at assignments 5 and 10, where we suspected conceptual difficulties were encountered, both measures fell. But on the next assignments, 6 and 11, the mean score continued to fall while the % of A's rose. This is consistent with the hypothesis that students encountered new conceptual difficulties in assignments 5 and 10. When the students first deal with new concepts, almost everyone struggles. The better students then master the concepts, and hence the percentage of A's recovers. The weaker students do not master the concepts, and their grades plummet as they get deeper into material they don't understand, causing the mean score to continue falling even though the percentage of high scores is rising. We have found that this pattern is a more secure way of recognizing conceptual difficulties. When this pattern matches the results from a pure speed analysis as above, our confidence is even higher.

Other issues can be seen in this data, of course. The behavior in assignment 7 (introducing complex numbers) is the reverse of the usual pattern for introducing new concepts. This may be caused by a mix of factors, including the fact that problem set 7 includes certain problems (such as adding complex numbers) are extremely easy, raising the mean scores relative to the other assignments. Another factor is that the deadline to drop without a "W" being recorded took place between the 6$^{th}$ and 7$^{th}$ assignment, leading to a decrease in the number of weak students completing the assignment as they dropped

the class. We would like to better understand this situation, but since there is only a single assignment on this topic, it is difficult to sort out different effects. Note that the key pattern identified above requires having more than one assignment addressing the concepts, so that variations between the assignments can be observed. What is more, even with multiple assignment controlling for extraneous effects (such as did a big football game cause students to spend less time on homework) is difficult. With only a single assignment, controlling for such extraneous factors becomes even less possible.

**Transfer**

Now that we have presumptively identified patterns that indicate students are struggling with new concepts, the next question is to determine if good performance on those assignments indicates understanding the concepts, and in particular if such understanding will then transfer to improved work in later courses. The natural approach here is to look at correlations between success on specific assignments in different courses. Unfortunately, determining whether a correlation is caused by transfer of knowledge from one course to another, as opposed to occurring because high scores on assessments in both courses are caused by some third factor (IQ, being a good worker, etc.) is tricky. There is a natural level of background correlation between any two assessments reflecting these common factors. To address this issue, we set a fairly high hurdle for counting a correlation as indicating transfer. A correlation matrix of all assignments in both Trigonometry and Calculus 2 is created. This correlation matrix is then processed using the technique of Agglomerative Nesting[5] to produce a tree diagram (dendrogram) that shows which assignments are most closely correlated to each other. When applied to assignments from two different courses, this process typically produces a tree with two distinct sections, one for each course. This is not surprising. It is very reasonable that assignments in one semester will be more closely correlated to assignments given in the same course and same semester than they are to assignments given a year later in a different course. Furthermore, the background correlation effects can reasonably be assumed to be larger between "nearby" assignments than distant assignments (it is reasonable to assume that student effort next week is more closely correlated to the effort they put in this week than it is to the effort they will put in next year). However, on occasion assignments will be more closely correlated to assignments in a different semester than they are to assignments in the same semester. When this happens, we will treat this as indicating transfer has taken place. When correlations with "distant" assignments are larger than correlations with "nearby" assignments, it is reasonable to suggest that the correlation is caused by some special feature of those two assignments that is important enough to produce larger effects than the background correlations that affect assessments taken at nearly the same time and in the same class.

We applied this technique to assessments in Trigonometry and assessments a year later in Calculus 2. Assessments were scored in two different fashions. The scores on each assignment (where 10 is perfect) were recorded and are listed as "hw." However, especially in Calculus 2, the fact that students could repeat problem sets as many times as they liked sometimes led to almost all students getting perfect scores, in which case the scores contained very little information. So for each assignment we also computed an

"inverse time to perfect" marked "it." This was 1 over the number of attempts it took to reach a perfect score. If a student never obtained a perfect score, the "it" for that assignment was set to 0. With units of inverse time, "it" is a measure of speed to achieve a perfect score. Furthermore, using inverses produces better results since speed need not fit the linear model implicit in computing correlations. For example, the difference between getting a perfect score in 1 attempt and in 3 attempts seems larger than the difference between getting a perfect score in 5 attempts and in 7 attempts. In computing linear correlations between number of attempts, these differences would be treated the same. But in taking inverse number of attempts, 1/1 and 1/3 are much farther apart than 1/5 and 1/7, in keeping with our intuitive sense. When we apply these techniques to the data from Trigonometry and Calculus 2, we get the results in Figure 3 below.

**Dendrogram of Trigonometry and Calculus Online Assessments**

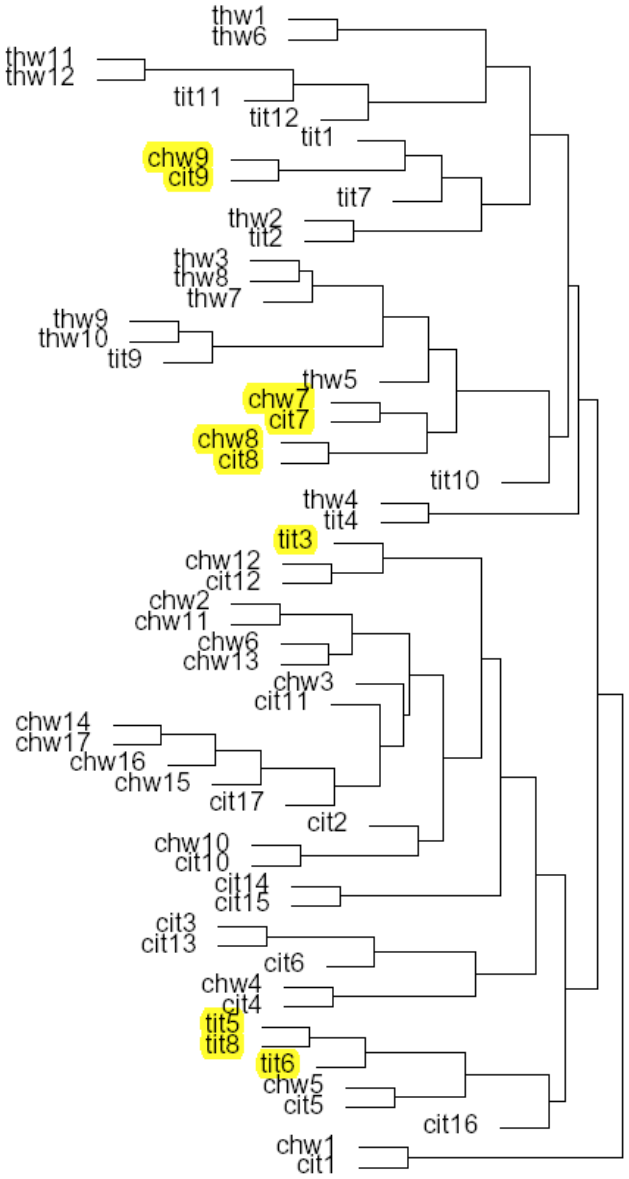


Figure 3

In the notation for this diagram, the first letter represents which class (Trigonometry or Calculus), the next two letters represent the type of measure, and the final number represents which assignment number. So for example, cit9 refers to the inverse times computed for the 9th assignment in Calculus 2.

Looking at this diagram, we see the top half of the tree primarily consists of trigonometry assessments, while the bottom half consists primarily of calculus assessments. This is in keeping with the assumption that assessments taken in the same semester will be more closely correlated than assessments taken a year earlier (or later) in a different class. However, there are a few assessments that are out of place, marked by highlighting in the diagram. In particular, the inverse times for the Trigonometry assignments 5 and 6 both appear in the Calculus section (along with those for assignments 3 and 7 as we will discuss below). Based on this procedure, it is reasonable to conclude that students who more quickly master material in Trigonometry that depends on the function concept (as assessed in assignments 5 and 6) will transfer their understanding of functions successfully to their later study of calculus. Note however that the overall score on the assignments is less important than the speed with which the material is mastered. This fits with the description of learning concepts at different levels from Van Hiele as previously noted.

The Calculus 2 assessments that appear in the Trigonometric half of the tree are the three assessments that deal with techniques of integration, where students frequently draw right triangles to determine which trigonometric substitution is appropriate, hence it is not surprising that they end up more closely correlated to Trigonometric skills. A review of the third assignment in Trigonometry (measured by tit3) showed there was a single problem that prefigured later work on symmetry of trigonometric functions, possibly explaining its appearance at the edge of the calculus section of the tree (though it is just at the boundary from being placed in the Trigonometric portion of the tree). Finally, the assignment marked tit8 referred to the complex variable assignment discussed earlier as assignment 7 which was difficult to classify. The reason for the reversed numbering is that students completing Calculus took trigonometry a year earlier, when the previous coordinator had switched the order of assignments 7 and 8.

It should be noted that the hurdle proposed to detect transfer is quite high. It is possible that transfer is taking place in terms of other concepts as well, but that the amount of transfer is not so large as to stand out relative to the level of background correlation. The technique used here is deliberately chosen to be quite conservative in indicating transfer, so that we only accept transfer as occurring when the evidence is very strong.

One final issue is the granularity of the developments noted here. It would be nice to have a measure that would tell us if individual students were ready for Calculus based on their work in Trigonometry. Unfortunately, the techniques we have described here are not suitable for such claims at present. Draper and Smith[6] recommend that before using statistical models for predicting performances of individuals, the F-statistic should be 4 times the critical value used for measuring statistical significance, and this data doesn't

reach that threshold. So the techniques as yet are only suitable for determining if how well a class is doing overall and not how well particular students are doing. As noted in the introduction though, since the classes in question range from 150-300 students, a measure of overall class performance is still useful from the standpoint of a course coordinator trying to manage a large lecture.

**Conclusions and Research Problems**

We began with three questions that have now been answered. It is possible to detect conceptual learning from analysis of procedural homework and it is possible to show that conceptual learning will transfer to later courses in at least some cases. The level of granularity possible at this stage is measurements of overall class performance and not of the individual student. To carry out this analysis, it is necessary that assignments be prepared so that there are several assessments over a concept in order that changes over different assessments can be compared. Looking at scores on individual assignments are not particularly meaningful, which is unsurprising since there are so many factors that can affect a single assignment.

In some ways, our work might seem unnecessary. Most instructors can recognize that students are likely to struggle with the function concept in trigonometry without a detailed analysis. But measuring exactly how they are struggling and how many are succeeding has traditionally been carried out using clinical interviews, which are not feasible for tracking a large class. Data-mining techniques that indicate the size of the bounce in the number of A papers on the second assignment gives us a way to automate the measurement of such learning in a fashion that scales easily to handle a large lecture. Using this technique, we can now address additional research questions.

One obvious research topic is whether we can extend this work to more courses, and especially for transfer from math to engineering, as opposed to just transfer between different math classes. This is currently difficult for several reasons. The hurdle we have proposed for identifying transfer is quite high and becomes higher the more remote the two courses are, hence showing transfer to engineering is difficult. Furthermore, we have found that differences between different semesters are much larger when looking at transfer to engineering than when looking at transfer between two math classes. We have had some success in measuring preparation for Linear Systems using acceleration of learning in Differential Equations, but the results have not been reproducible over different semesters. Our working hypothesis is that the instructor (which varies according to a rotation) is a significant variable in transfer between disciplines. Our reasoning is that math instructors are likely to have similar attitudes toward mathematical topics that transfer to later math courses, but may have more varied attitudes toward more applied materials. This hypothesis seems consistent with the limited data we have, but we are waiting to accumulate sufficient information for each instructor to test this hypothesis properly.

Another obvious extension is to consider the specific sorts of errors students make in trying to analyze conceptual learning. The difficulty with this approach is that students

are too inventive in the types of errors they make. Furthermore, they often make multiple errors in a single problem. We have thus far been unable to develop an artificial intelligence system that is capable of classifying a sufficient number of student errors to support a proper data-mining procedure.

Finally, we hope to move this research into practice. Currently, the feedback loop for teaching is so long that it is rarely completed. The instructor lectures, but homework isn't turned in until a week later and often not graded until a week after that. Analysis of how students are doing is anecdotal and doesn't come soon enough to enable the instructor to make changes when things aren't going well. The long-term goal is to prepare a system that provides real-time feedback to the coordinator so that each week or so the system provides an email indicating roughly what percentage of the students are understanding the key concepts and are likely to be successful in applying these ideas in future classes. With prompt feedback, the instructor may be able to make adjustments on the fly to address situations where the class is not understanding in time to help students before they get left behind. The research in this paper suggests such a system is feasible as online homework systems become more sophisticated.

**Bibliography**


1. http://www.webassign.net/info/users.html (accessed Jan. 13, 2007).
2. http://nces.ed.gov/nationsreportcard/ (accessed Jan. 16, 2007).
3. http://nces.ed.gov/timss/ (accessed Jan. 16, 2007).
4. Van Hiele, P., *Structure and Insight*, 1986, Academic Press, Orlando, FL.
5. Kaufman, L. and P. Rousseeuw, *Finding Groups in Data*, 1990, Wiley, New York, NY.
6. Draper, N. and H. Smith, *Applied Regression Analysis*, 1981, Wiley, New York, NY.