

Gathering and Timely use of Feedback from Individualized On-line Work

Matthew Hall¹, Joyce Parker², Berhouz Minaei-Bigdoli³, Guy Albertelli⁴, Gerd Kortemeyer⁵ and Edwin Kashy⁶

Abstract - Technology has enabled instructors to efficiently create and distribute a wide variety of educational materials, assignments, assessments, etc. These include numerous types of formative conceptual and algorithmic exercises for which prompt feedback and assistance can be provided to students as they work on assigned tasks. At the same time, the technology records and dynamically organizes a vast amount of information on students' interaction with and understanding of these materials. We present recent developments that allow rapid interpretation of such data in identifying students' misconceptions and other areas of difficulty, so that concurrent or timely corrective action can be taken. This information also facilitates detailed studies of the educational resources used and can lead to redesign of both the materials and the course.

Index Terms - Feedback, computer-graded assignments, detecting and addressing misconceptions, formative and summative assessment.

INTRODUCTION

Information technology has become common in higher education, as reflected in numerous course management systems. The ubiquitous web browser represents a remarkable enabling tool to get information to and from students. For the student, that information can be textual and illustrated, not unlike that presented in a textbook, but also include various simulations representing a modeling of phenomena, essentially experiments on the computer. Its greatest use however is in transmitting information as to the correct or incorrect solutions of various assigned exercises and problems. It also transmits guidance or hints related to the material, sometimes also to the particular submission by a student, and provides the means of communication with fellow students and teaching staff.

While several meta-analyses of the effects of assessment with immediate feedback to the student on their learning are positive [1,2], the range of effect size is considerable [3], and can even be negative [2,4,5,6]. Even within our own model systems CAPA, LectureOnline, and LON-CAPA, when used just for homework, a range of partly contradictory observations were made [7,8]. There will not be a general

answer to the question of whether or not systems like LON-CAPA are beneficial - after all, they are just tools, not a curriculum. Instead, effectiveness will depend on how they are used, and with which material. There is no doubt however that timely feedback, as discussed in this paper, is crucial for ensuring effective use.

Course management systems can and often do record all information transmitted to and from the student. That large amount of data, especially in large courses, is much too dilute for the faculty to interpret and use without considerable pre-processing. We discuss functions that make that vast amount of data useful in a timely fashion. The instructor can then give students useful feedback, either promptly enough that student can benefit while still working on current task, or at a later date to clarify misconceptions and address lack of understanding. A preliminary report on some of this work was presented at the 2002 meeting of the FIE conference as a work-in-progress contribution [9].

THE TOOL

The system we use is LON-CAPA, (Learning Online Network with a Computer-Assisted Personalized Approach) [10]. This system, while similar to many others in most aspects, differs in three important ways relevant to the current discussion. The first is its capability to individualize problems, both algorithmic numerical exercises as well as problems that are qualitative and conceptual so that numbers, options, images, etc. differ from student to student. [11]. The second is in the tools provided that allow instructor to collaborate in the creation and sharing of content in a fast and efficient manner, both within and across institutions, thus implementing the initial goals of the WWW [12]. And the third is its one-source multiple target capabilities: that is, its ability to automatically transform one educational resource, for example a numerical or conceptual homework question, into a format suitable for multiple uses. The same source code which is used to present problems for on-line homework can also generate them for an on-line examination or for a printed version suitable for a proctored bubble sheet examination which is later machine scored [13]. LON-CAPA provides a superset of the functionalities of the two well-tested previous systems, CAPA and LectureOnline from which it originated [14,15]. A

¹ Matthew Hall, Michigan State University, Division of Science and Math Education, hallmat3@msu.edu

² Joyce Parker, Michigan State University, Division of Science and Math Education, parker13@msu.edu

³ Berhouz Minaei-Bigdoli, Michigan State University, Department of Computer Science, minaeibi@cse.msu.edu

⁴ Guy Albertelli, Michigan State University, Division of Science and Math Education, albertel@msu.edu

⁵ Gerd Kortemeyer, Michigan State University, Lyman Briggs School and Division of Science and Math Education, korte@lite.msu.edu

⁶ Edwin Kashy, Michigan State University, Department of Physics and Astronomy and NSCL, kashy@nscl.msu.edu

summary of performance results obtained this past decade using CAPA for homework, quizzes, and summative as well as formative examinations has been published [15].

DISCUSSION

The simplest function of the statistics tools in the system is to quickly identify areas of student difficulties. This is done by looking at the number of submissions students require in reaching a correct answer, and is especially useful early after an assignment is given. A high degree of failure indicates the need for more discussion of the topic before the due date, especially since early responders are often the more dedicated and capable students in a course. Sometimes a high degree of failure has been the result of some ambiguity in wording or, mostly in newly authored problem resources, the result of errors in their code. Difficulty is then 'sky high'. Quick detection allows correction of the resource, often before most students have begun the assignment. Figure 1 below shows a plot of the ratio of number of submissions to number of correct responses for 17 problems, from a weekly assignment before it was due. About 15% of the 400 students in an introductory physics course had submitted part or most of their assignment.

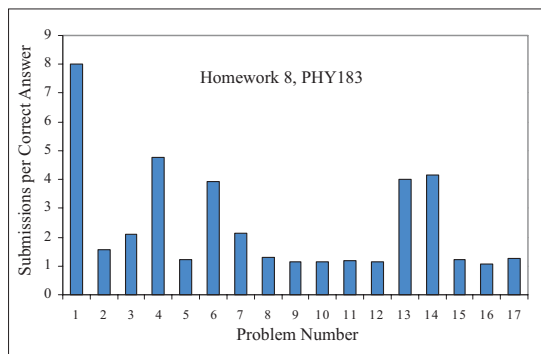


FIGURE 1
ONE EARLY MEASURE OF A DEGREE OF DIFFICULTY

The data of Figure 1 is also available as a table which also lists the number of students who have submissions on each problem. Figure 1 shows that five of the questions are rather challenging, each requiring more than 4 submissions per success on average. Problem 1 requires a double integral in polar coordinates to calculate a center of mass. Problem 14 is a qualitative conceptual question with six parts and it becomes likely that one part or another will be missed. Note that incorrect use of a calculator or incorrect order of operation in a formula would not be detected in Figure 1 because of their relatively low occurrence. Note also that an error in the unit of the answer or in the formatting of an answer is not counted as a submission. In those instances, students re-enter their data with proper format and units, an important skill that students soon acquire without penalty.

An important task of the feedback tools for the instructor is to help identify the source of difficulty in numerical algorithmic questions and, for qualitative questions, the misconceptions students have on the topic. We first look at a

numerical problem designed to check that students use their electronic calculators properly. The problem is simple enough: Use your calculator to take the square root of 10^{-27} . (The exponent for some students is -21 , or -23 , etc...). Students' submissions show a large number of 3.16×10^{-n} . The reason is that while students use their calculator properly to take the square root of 4.5×10^{-27} , many enter $10E-27$ when the number is just 10^{-27} . Here the computer feedback forces students to become aware that something is wrong and they often seek help as they do not see how they can be wrong. Other questions test that calculators are properly set for trigonometric functions (radians, degrees), or that the order of operations is properly used as in the fraction 'A' divided by 'BC', where 'A', 'B', and 'C' are randomized numerical values. Once particular errors are detected it is possible to include specific hints which will be displayed when that error is detected. This automates the feedback process but requires considerable work to implement.

Student responses to two qualitative exercises, one from physics and the second essentially vector math, illustrate the way that the analysis tool detects difficulties and their source, specific misconceptions. The physics question is problem 14 from assignment 8, which as indicated above, had five days before it was due. As shown in Figure 1 that problem averaged at that time slightly more than 4 submissions per successful solution. There were 50 correct solutions as a result of 208 submissions by 74 students. The order in which the six statements are presented varies among students. Each statement is selected randomly from one of the six concept groups. Each concept group focuses on a particular aspect in the question. Success rate on each concept for the initial submission is shown in Figure 2.

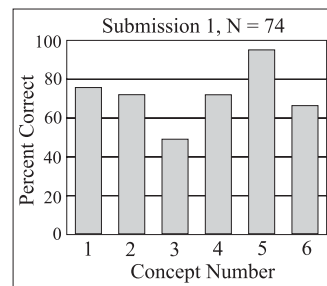


FIGURE 2
SUCCESS (%) IN INITIAL SUBMISSION FOR SELECTING THE CORRECT ANSWER TO EACH OF SIX 'CONCEPT' STATEMENTS

While concept '3' is quite clearly the most misunderstood, there is also a large error rate for concepts '2', '4' and '6'. About one third of the students succeeded on their first submission for all six concepts groups and thus earned credit on their first submission. This can be seen by looking at the decreasing number of submissions from Figure 2 to Figure 3. Note the pattern in the initial submissions persists in subsequent submissions with only minor changes.

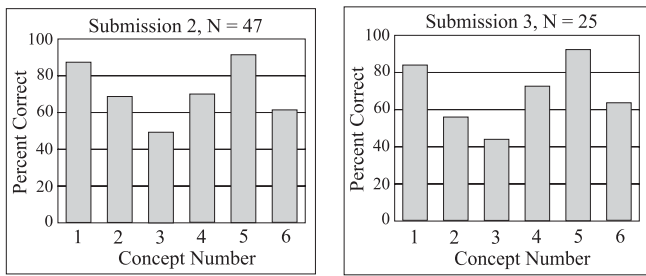


FIGURE 3

SUCCESS RATE ON SECOND AND THIRD SUBMISSIONS FOR ANSWERS TO EACH OF SIX 'CONCEPT' STATEMENTS

The text of the problem corresponding to the data in Figures 2 and 3 is shown in Figure 4.

A frictionless pulley with mass M_a is attached to the ceiling, in a gravity field g . Mass M_c is smaller than mass M_b . The tensions T_x , T_y , T_z , and the constant g are magnitudes. For each statement, select a response.

Choices: **Greater than, Less than, Equal to, True, False.**

- T_z is T_y .
- $(M_b + M_c + M_a)g$ is T_x .
- T_y is $(M_b)g$.
- The magnitude of the acceleration of M_b is that of M_c .
- The center-of-mass of M_b , M_c , and M_a accelerates.
- T_x is $T_z + T_y$.

FIGURE 4

RANDOMLY LABELED CONCEPTUAL PHYSICS PROBLEM

The labels in the problem are randomly permuted. In the version of the problem shown in Figure 4 the first question is to compare tension T_z to T_y . It is the most commonly missed statement, corresponding to concept '3' of Figures 2 and 3. The incorrect answer given by over 90% of the students is that the two tensions are equal, which would be the answer for a pulley with negligible mass. That had been the case in an assignment two weeks earlier. This error was addressed by discussion in lecture and by a demonstration showing the motion for a massive pulley with unequal masses. This quickly impacted the subsequent response pattern. Note that solutions to the versions of the problems use as illustrations are given at the end of this section.

The next example is shown in Figure 5. It deals with the addition of two vectors. The vectors represent the possible orientations and rowing speed of a boat and the velocity of water. Here also the labeling is randomized so both the image and the text vary for different students. Students are encouraged to discuss and collaborate, but cannot simply copy from each other.

A river is to be crossed by a boy using a row boat.

Assume that the water has uniform velocity, represented above by the vector labeled B. The rowing speed of the boy and a set of possible orientation of his boat are also shown. Select an answer for each below.

Choices: **Greater than, Less than, Equal to.**

- Time to row across for A is for H
- Time to row across for K is for C
- The distance traveled in crossing for H is for K
- Time to row across for K is for H
- For an observer on shore, the speed of the boat for K is for C

FIGURE 5

VECTOR ADDITION CONCEPT PROBLEM

The question has 14 statements, 7 of which are simply due to reversing the two quantities being compared. Each student sees five statements (in multiple formats because of random labeling) one from each of the five concepts. These are, stated rather cryptically,

- Concept 1 - Equal transverse velocities, equal time to cross.
- Concept 2 - Greater transverse velocities, shorter time to cross. Boat is directed downriver.
- Concept 3 - Greater transverse velocities, shorter time to cross. Boat directed upriver.
- Concept 4 - Distance traveled while crossing; less: if total velocity more normal to flow direction, more: if along the flow direction.
- Concept 5 - The more aligned two vectors are, the greater is their sum.

The upper graphic of Figure 6 shows once again the success rate of 350 students on their initial submission, but this time in more detail showing all the possible statements.

There are two variations for the first three concepts and four for the last two.

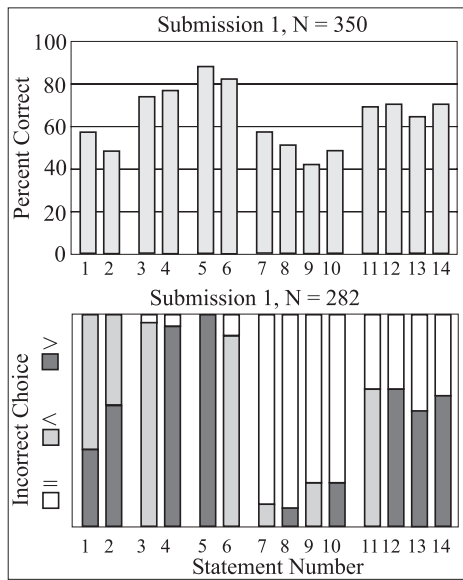


FIGURE 6

UPPER SECTION: SUCCESS RATE FOR EACH POSSIBLE STATEMENT.
 LOWER SECTION: RELATIVE DISTRIBUTION OF INCORRECT CHOICES, WITH DARK GRAY AS “GREATER THAN”, LIGHT GRAY AS “LESS THAN” AND CLEAR AS “EQUAL TO”

The lower graph in Figure 6 illustrates the distribution of incorrect choices for the 282 students who did not get earn credit for the problem on their first submission. The stacked bars show the way each statement was answered incorrectly. This data gives support to the ‘concept group’ method, not only in the degree of difficulty within a group as reflected by the Percent Correct in Figure 6, but also by the consistency of the misconception as seen from the Incorrect Choice distribution. Statements 3 and 4 in Figure 6 present ‘Concept 2’, that greater transverse velocities result in a shorter crossing time, with the vectors in reverse order. Statement 3 reads ‘Time to row across for K is for C’, and statement 4 is ‘Time to row across for C is for K’. Inspection of the graph indicates the students made the same error, assuming the time to row across for K is less than the time to row across for C, regardless of the manner in which the question was asked. Few students believed the quantities to be equal. In concept group 3, statements 7, 8, 9 and 10, “equal to” is predominantly selected instead of ‘greater than’ or ‘less than’ as appropriate. This detailed feedback makes it easier for the instructor to provide help so that students discover their misconceptions. Finally, as in the previously discussed numerical example, particular hints can be displayed, triggered by the response selected for a statement or by triggered by a combination of responses for several statements.

We now turn to the task of evaluating resources used in examinations. Examinations as assessment are most useful when the content includes a range of difficulty from fairly basic to rather challenging problems. An individual problem within an examination can be given a difficulty index (DIFF) simply by examining the class performance on that problem. It

equals 1 minus the ratio of earned points on the problem to possible points. If all students get full credit, $DIFF = 0$, and if no students earn any points, $DIFF = 1$. Table 1 below shows an analysis for the first two mid-term examinations in Spring 2004.

TABLE 1
 ANALYSIS OF EXAMINATION PROBLEMS (N=393)
 DIFF = DIFFICULTY INDEX DISC= DISCRIMINATION INDEX

Problem Number	DIFF Exam 1	DISC Exam 1	DIFF Exam 2	DISC Exam 2
1	0.20	0.40	0.70	0.24
2	0.16	0.31	0.13	0.20
3	0.40	0.40	0.19	0.31
4	0.44	0.57	0.41	0.57
5	0.32	0.38	0.52	0.11
6	0.00	0.00	0.18	0.26
7	0.23	0.33	0.70	0.36
8	0.21	0.24	0.57	0.35
9	0.36	0.63	0.55	0.58
10	0.40	0.59	0.87	0.14
11	0.25	0.31		

A second measure of a problem’s usefulness in assessing performance is its discrimination index (DISC), which is also shown in Table 1. It is derived by comparing how students whose performance places them in the top quartile of the class score on that problem compared to those in the bottom quartile. The particular problem is excluded in making that comparison. To compute DISC, students are ranked based on their performances on the other problems in an examination, or group of examinations. The ratio, points earned to points possible, by the upper quartile minus that same ratio for the lower quartile give the value of DISC. This leads to possible values from -1 to $+1$. A negative value means that students in the lower quartile scored better on that problem than those in the upper, surely not a good recommendation for that problem. A value close to $+1$ indicates the problem was difficult for the students in the lower quartile but was answered correctly by most students in the upper quartile. We can see that Exam 1 was on the average somewhat less difficult than Exam 2. Problem 10 in Exam 2 has $DISC=0.14$ and $DIFF=0.87$, indicating it was difficult for all students. The students did not understand the concepts involved well enough to differentiate this problem from a similar problem they had seen earlier. In Exam 1, problems 3, 4, 9, and 10 are not too difficult and nicely discriminating.

One striking entry in Table 1 is for problem 6 in Exam 1. There both DIFF and DISC are 0. No difficulty and no discrimination! As the reader has probably guessed this is a result of a faulty problem. While the error had been detected in the process of preparing the examination and corrected, the correction had not been propagated back into the exam. This unfortunate human error was addressed to the joy of the students by giving all students full credit on the problem. Also, a request was submitted to modify LON-CAPA so that in the future an instructor will be warned of such a circumstance.

The distribution of scores on homework assignments differs considerably from that on examinations. This is clearly seen in Figure 7.

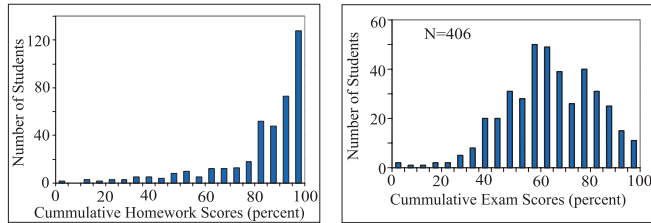


FIGURE 7
GRADES ON THE FIRST SEVEN HOMEWORK ASSIGNMENTS AND ON THE FIRST TWO MIDTERM EXAMINATIONS

The correlation of homework and examinations is moderate ($r=0.43$). Students with a good exam a score tend to score high on homework but the reverse is not as true. This can be seen in the 3-D plot of the Figure 7 data in Figure 8. Homework grades peak near 100% as motivated students tend to repeat problems until a correct solution is obtained. Other students just seek a formula to plug in their values and ‘get the points’. Students can also seek help in a learning center, staffed by both graduate and undergraduate teaching assistants. Unfortunately a significant number of motivated students are also not well prepared for the course, especially in math. They can get help, but observations in the help room show that many do not grasp the underlying concepts that lead to the solution.

Students also often interpret a high homework grade as indication that they are doing well in the course. To counter that misconception, a readily accessible on-line grade extrapolator provides students a review of their performance to date in the various components of the class, quizzes, mid term exams, and homework. They enter their own estimate of their future performance for the remainder of the semester, as well as for the final examination. This tool then projects a final grade, thus keeping students aware of their progress.

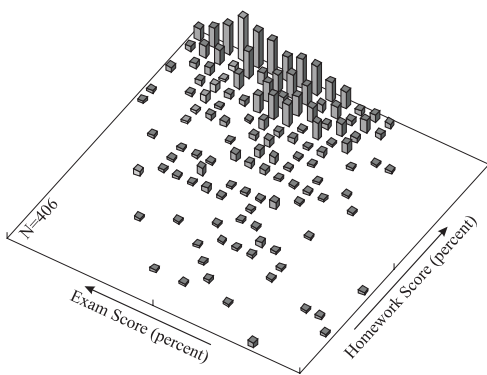


FIGURE 8
HOMEWORK VS. EXAM SCORES. THE HIGHEST BIN HAS 18 STUDENTS.

Another method of providing feedback is through numerous short quizzes during the lecture periods. A problem whose solution seemed rather simple when solved by the instructor on the overhead can look quite different when part

of a quiz, and thus can provide a wake up call about the students’ understanding. These individualized quizzes would also require too much time to distribute were it not possible to print quizzes in “anonymous” mode, i.e., with students getting any paper and entering a corresponding code on their scoring form so that the test can be graded as each test paper differs.

Feedback is also given to students in response to work requiring them to compose and submit an essay. The essays are later read and assessed by the teaching staff. This old-fashioned subjective reading and evaluation is still king, but the task is considerably easier with the technology. Electronic submissions by students within the system, automatic grade recording, and feedback to the student via the system are the main benefits. Next to face-to-face interaction with students, these essays remain one of the best ways to assess students’ understanding.

Finally, as a result of feedback on students’ work, those doing very poorly can be identified quite early. [16,17]. The important question is what to do then. One colleague contacted 50 such students early, invited them to come and discuss their difficulties, and offered to help. One student appears to have benefited significantly, and that is indeed better than none [17]. This brings to mind the following tale: A man walking on the beach observes a woman bending down and tossing something into the ocean, over and over. As he approaches, he notices that she is picking up a starfish each time and tossing it into the water. With starfishes strewn all over the beach he addresses her saying “This can’t possibly make a difference!”. The woman bends down, picks up a starfish, tosses it into the ocean. “To him it did.”, she says [18].

Solutions to illustrative problems [19]:

- Figure 4: 1-less, 2-greater, 3-less, 4-equal, 5-true, 6-greater.
- Figure 5: 1-less, 2-greater, 3-less, 4-equal, 5-greater.

CONCLUSIONS

Technology does indeed provide means to get considerable feedback on many aspects of teaching and learning. To make good use of that feedback is a far greater challenge. We have been using LON-CAPA for both formative and summative assessment. Our ability to detect, to understand, and to address student difficulties is highly dependent on the capabilities of the tool. Feedback from numerous sources has considerably improved the educational materials, which is a continuing task. The software in its current state is not only quite functional but also relatively easy to use, a far cry from its form in its early years [20].

ACKNOWLEDGMENT

We thank the National Science Foundation for grants supporting this work through the Information Technology Research (ITR 0085921) and the Assessment of Student Achievement (ASA 0243126) programs. Support in earlier years of this project was also received from the Alfred P Sloan and from the Andrew W Mellon Foundations. We are grateful to our own institution, Michigan State University and to its

administrators for over a decade of encouragement and support. Discussions with our colleagues D. A. Kashy, M. Thoennessen, F. Berryman, and S. Raeburn were particularly useful.

REFERENCES

- [1] Azevedo, R, Bernard, R, M, "A Meta-analysis of the Effects of Feedback in Computer-based Instruction", J. Educational Computing Research 13, 111-127. (1995).
- [2] Mason, B, J, Bruning, R, "Providing Feedback in Computer-based instruction: What the Research Tells Us." <http://dwb.unl.edu/Edit/MB/MasonBruning.html> (2003).
- [3] Bonham, S, W, Beichner, R, J, and Deardorff, D. L., "Online Homework: Does It Make a Difference?" The Physics Teacher (2001)
- [4] Bransford, J, D, Brown, A,L, & Cocking, R, R, (Eds). "How People Learn: Brain, Mind, Experience, and School", Washington, D.C.: National Academy Press, (2000).
- [5] Kluger, A, N, DeNisi, A, "The Effects of Feedback Interventions on Performance: Historical Review, a Meta-Analysis and a Preliminary Feedback Intervention Theory". Psychological Bulletin, 119, 254-284 (1996).
- [6] Kluger, A, N, DeNisi, A, "Feedback interventions: Toward the understanding of a double-edged sword", Current Directions in Psychological Science, 7, 67 (1998)
- [7] Pascarella, A, M, "The Influence of Web-Based Homework on Quantitative Problem-Solving in a University Physics Class", NARST Annual Meeting Proceedings (2004)
- [8] Kotas, P, "Homework Behavior in an Introductory Physics Course", Masters Thesis (Physics), Central Michigan University (2000)
- [9] Albertelli, G, Minaei-Bidgoli, B, Punch, W, F, Kortemeyer, G, and Kashy, E "Concept Feedback in Computer-Graded Assignments" IEEE Frontiers in Education Conference Proceedings, vol. 32, (2002)
- [10] Kortemeyer, G, Bauer, W, Kashy, D,A, Kashy, E, and Speier, C, "The LearningOnline Network with CAPA Initiative", IEEE Frontiers in Education Conference Proceedings, vol. 31,(2001) p. 1003. See also www.lon-capa.org. The site contains information on the system as well as links to numerous presentations and publications.
- [11] Kashy, E, Gaff, S, J, Pawley, N, H, Stretch, W, L., Wolfe, S, L., Morrissey, D.J., Tsai, Y., "Conceptual Questions in Computer-Assisted Assignments", American Journal of Physics, Vol, No 63, (1995), pp. 1000-1005.
- [12] See <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/Proposal.html> and also <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/DesignIssues/Multiuser.html>
- [13] Albertelli, G, Kortemeyer, G, Sakharuk , A, and Kashy, E, "Personalized Examinations in Large On-Campus Classes", IEEE Frontiers in Education Conference Proceedings, vol. 33, (2003)
- [14] Kortemeyer,G, and Bauer, W, "Multimedia Collaborative Content Creation (mc3) - The MSU LectureOnline System", Journal of Engineering Education, Vol. 88 No. 4, (1999) p. 405
- [15] Kashy, D A, Albertelli, G, Kashy, E, and Thoennessen, M, "Teaching with ALN Technology: Benefits and Costs", Journal of Engineering Education, Vol. 90, No. 4, October 2001, pp. 499-505
- [16] Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer G., Punch, W.F., "Predicting Student Performance: An Application of Data Mining Methods with an educational Web-based System", (IEEE/ASEE) FIE 2003 Frontier In Education, Nov. 2003 Boulder
- [17] Thoennessen, M, Kashy, E, Tsai, Y, Davix, N, Ne, "Impact of Asynchronous Learning Networks in Large Lecture Classes", Group Discussion and Negotiation Vol. 8 (1999) pp 371-384 and private communication. Devoting more time and resources to help these students would perhaps have yielded better results.
- [18] The story can be found in various forms in many places including several internet sites. The original source is unknown.
- [19] Illustrative examples were written by one of us (EK) and are from Michigan State University's LON-CAPA physics problem collection.
- [20] Moore, Geoffrey, A, "Crossing the Chasm", HarperCollins, 2002. Users of LON-CAPA in its very early beta versions would appreciate the statement on p.31, as these early adopters also forgave "...ghastly documentation..." as well as "... bizarrely obtuse methods of invoking needed functions ..."