On the effectiveness of active-engagement microcomputer-based laboratories

Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg
Department of Physics, University of Maryland, College Park, MD 20742

One hour active-engagement tutorials using microcomputer based laboratory (MBL) equipment were substituted for traditional problem-solving recitations in introductory calculus-based mechanics classes for engineering students at the University of Maryland. The results of two specific tutorials, one on the concept of instantaneous velocity and one on Newton's third law were probed by using standard multiple-choice questions and a free-response final exam question. A comparison of the results of eleven lecture classes taught by six different teachers with and without tutorials shows that the MBL tutorials resulted in a significant improvement compared to the traditional recitations when measured by carefuly designed multiple choice problems. The free-response question showed that, although the tutorial students did somewhat better in recognizing and applying the concepts, there is still room for improvement..

I. INTRODUCTION

It is by now well documented that students in introductory university physics have considerable difficulties with the fundamental concepts of Newtonian mechanics.[1] The computer is often cited as a panacea for solving educational problems, but anecdotal evidence on the use of computers in a variety of situations suggests that the results may not be uniformly satisfactory. In this paper, we consider whether computer activities, when based on results of physics education research and designed following principles from models of cognition and learning, can successfully teach basic physics concepts to a large fraction of students in an introductory physics class, without a large investment in time or equipment.

The specific issues we investigate are the learning of the concepts of instantaneous velocity and Newton’s third law. Facility with these concepts is essential to an understanding of mechanics and addresses general issues (such as the relation between a quantity and its rate of change and the nature of interactions) that play an important role throughout introductory physics. These concepts are known to be difficult for many students. We target each of these difficulties with one hour of active-engagement microcomputer-based laboratory (MBL) activities.[2]

Of the previous work on this subject, the most relevant is the oft-cited paper of Thornton and Sokoloff.[3] They report that introductory physics students' understanding of velocity graphs could be significantly improved using an MBL curriculum they developed. They evaluated the effect of their curriculum using a set of multiple-choice velocity questions (VQ) in which students were required to match a description of a motion to a velocity graph. They then demonstrated that students who were given four hours of group-learning guided-discovery active-engagement MBL proved significantly more successful in choosing the correct graphs than those who only received traditional instruction.

The results are dramatic, with a large fraction of the students missing all but the simplest of the five velocity graph questions after traditional instruction.[4] After the MBL activities, the error rate drops to below 10% on all the questions. This result is strikingly robust and has now been confirmed at dozens of universities and colleges.[5] In addition to confirming the difficulty reported by other researchers, a difficulty that many instructors find surprising, they demonstrate the existence of a solution. This work is often cited as an indication that interactive-engagement MBL activities are highly effective. Several questions remain to be addressed, however.

Q1. Is the improvement due to the MBL activity or to the extra time spent on the topic?

Q2. How functional is the improved knowledge? Does a significant improvement on multiple choice questions imply that the students can use these concepts in other contexts such as problem solving?

Q3. Can other non-MBL activities be equally effective in producing improved learning?

This study explores the first two of these three questions and touches briefly on the third. We encourage others to address the third.

II. CLASS ENVIRONMENT

The study was performed in the calculus-based introductory physics class for engineering students at the University of Maryland in College Park. This is a three semester sequence where traditionally each course in the sequence has three hours of lecture from a faculty member and one hour of recitation from a teaching assistant per week. The second and third semesters have an associated two hour laboratory which is run independently. This study involves the first semester course which covers Newtonian mechanics.

To allow students to have more interaction with faculty, lecture classes are formed of 50-150 students with each class taught by a single faculty member. Each lecture class is divided into sections of about 25 for recitations and laboratories. The textbook[6] and the approximate outline of the course content are chosen by a course committee, but otherwise, each faculty member acts independently. There are no common exams and there is no laboratory component in the mechanics course. The lecture hours tend to be traditional with little student interaction. Occasionally, faculty distribute in-class worksheets or engage the class with questions and discussion, but this is rare. The recitation hour typically consists of a graduate student solving problems at the board. Often there is a brief quiz (usually one of the homework problems) and sometimes the choice of problem discussed is based on student questions or requests. Teaching assistants typically receive no special training for these sessions.

We suspected that our traditional lecture plus recitation environment suffered the oft-reported problems of teaching mechanics: students appear to master algorithmic problem solving techniques but fail to make significant improvement in their understanding of the fundamental concepts.[7] To try to improve this situation, we introduced an experimental research-based instructional technique which we refer to as *tutorials*. This method was developed by Lillian McDermott and the Physics Education Group at the University of Washington to improve student understanding of fundamental physics concepts in a cost-effective manner within the traditional lecture structure.[8]

These tutorials have the following components:

1. A 10 minute ungraded "pretest" is given in lecture once a week. This test asks qualitative conceptual questions about the subject to be covered in tutorial the following week.
2. The teaching assistants and faculty involved participate in a 1.5 hour weekly training session.
3. A one hour (50 minute) tutorial session replaces the traditional problem-solving recitation. Students work in groups of three or four and answer questions on a worksheet that walks them through building qualitative reasoning on a fundamental concept. At least two teaching assistants serve as facilitators, asking leading questions in a semi-Socratic dialog[9] to help the students work through difficulties in their own thinking.
4. Students have a brief qualitative homework assignment in which they explain their reasoning. This is in part of their weekly homework which also includes problems assigned from the text.
5. A question emphasizing material from tutorials is asked on each examination.

At the University of Washington, tutorial worksheets are developed over a period of many years through an iterative cycle of research/curriculum-development/instruction. They often use "cognitive conflict". In this approach, situations are presented which trigger the common student difficulties revealed by research. The facilitators then help those students who show the predicted difficulties work through their ideas themselves. McDermott refers to this process as *elicit/confront/resolve*.[10] Since the fall semester of 1993, we have implemented many of these tutorials at the University of Maryland in one or more lecture classes each semester. We supplemented them by a number of tutorials we developed ourselves in the same framework. Since we had a laboratory-style room with computers available, some of our tutorials were based on MBL activities or on simulations.

III. ACTIVITIES

We created two tutorials using MBL activities, one to assist students with the concept of instantaneous velocity and one to help them with Newton's third law. Our MBL equipment used a computer connected to a universal laboratory interface box (ULI) with a sonic ranger and two force probes. [11]

The first tutorial was based directly on the MBL activities developed by Thornton and Sokoloff labs in *Tools for Scientific Thinking*.[12] We extracted from their velocity labs what we considered the essential elements, following the guidance in their paper (ref 3). In the tutorial, students walk in front of a sonic ranger which provides immediate feedback and reduces data-collection drudgery. In the tutorial, students use their own bodies to

- familiarize themselves with the equipment by creating a series of position graphs;
- create a series of simple velocity graphs;
- match a given complex velocity graph.[13]

In each case, the students work together in groups of three or four. They discuss and make predictions of what the graph will look like or how they have to move in order to produce the desired result and they write these predictions on their worksheets. The entire activity is easily completed in one fifty-minute period.

The second tutorial is based on suggestions of Laws, Thornton and Sokoloff.[14] Newton's third law is explored by having students connect the force probes to two low-friction carts and observe the result of their interaction. The apparatus is sketched in Fig. 1.

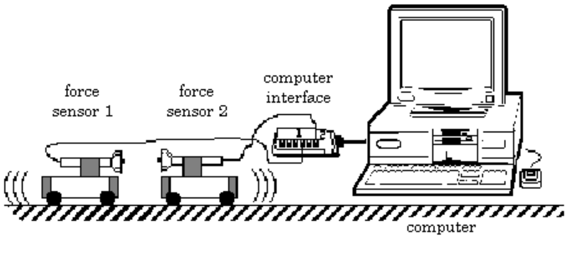


Fig. 1: The arrangement for the Newton 3 tutorial.

In the tutorial, students

- psychologically calibrate the force probe by pushing and pulling on it and watching the result on the computer screen;
- predict the relative size of forces for a light car pushing a heavy truck;
- predict and observe the forces two identical carts exert on each other when one pushes the other;
- predict and observe the forces two carts exert on each other when one is weighted with iron blocks;
- predict and observe the forces two identical carts exert on each other when one collides with the other;
- predict and observe the forces two carts exert on each other when one collides with a second weighted with iron blocks.

In addition, the students are asked to draw free body diagrams and use them it their predictions. Again, this activity is easily completed in one fifty-minute period.

IV. EVALUATION TOOLS

In order to evaluate the success of our interventions, we must decide what we mean by "success". This will play an important role in determining our approach to evaluation. What we mean by success is, in turn, determined by our model of student understanding and learning.[15] The critical element of this model for our application is that a student may "have" an item of knowledge, that is, be able to recall it in response to a narrow range of triggers, but be unable to recall and apply it in a wide range of appropriate circumstances. We want our evaluations to test for robust functional understanding.

Four plausible and frequently used approaches to evaluation are:

1. Measure student and faculty satisfaction with a survey or questionnaire.
2. Measure student learning using a multiple-choice test designed using the results of physics education research on commonly found errors to specify attractive distractors.
3. Measure student learning using long-answer exam questions -- problems or open-

expression questions in which students explain and discuss their answers.

4. Measure student learning through recorded problem interviews.

The first approach is the simplest and most commonly used, but although both student and faculty satisfaction is important in motivating student work, and presumably therefore student success, the link between satisfaction and learning is highly indirect. Indeed, students whose primary goal is a good grade may find higher satisfaction in a course that produces a good grade without improved learning, since improved learning often requires time and painful effort. We do not expect this measure to correlate well with functional understanding.

The second approach is easy to deliver, but requires a substantial effort to develop. The results can be highly suggestive, but multiple choice tests can be difficult to interpret. They have a tendency to overestimate the student’s learning since they can sometimes be answered correctly by means of incorrect reasoning[16] or by "triggered" responses that fail to represent functional understanding. On the other hand, the use of common misconceptions as distractors produces "attractive nuisances" that challenges the students' understanding. Students that get the correct answer despite this challenge are likely to having a good understanding of the topic in question. We expect therefore that this approach does give some indication of the robustness of a student’s possession of and confidence in a correct answer.

The third approach is easy to deliver, but the analysis can be time consuming. Student answers must be read in detail and classified by the understanding displayed. The functionality of student knowledge is rather well-tested by this approach since the student is being asked to produce the desired knowledge within the context of a problem and without the most common and automatic triggers. It has the defect that students occasionally give answers too incomplete or ambiguous to let us see what they are thinking.

The fourth approach is the most effective since it permits the researcher to observe in detail the functionality of the student’s knowledge by the presentation of a variety of contexts. The researcher can follow up suggestive responses with more detailed and individually designed questions, but it is highly time consuming. In addition to the recording time (usually one or more hours per student), the recordings must be transcribed and analyzed. This approach is thus impractical for evaluating the distribution of student knowledge throughout a large class.

We have therefore chosen to combine the second and third approaches. We use as our primary evaluation tool the multiple-choice velocity questions (VQ) from Thornton and Sokoloff and the Force Concept Inventory (FCI) of Hestenes et al.[17] We supplement these with a long-answer examination question to provide a probe of the functionality of the students' knowledge in a more complex context.

In order to permit a comparison of our students with other classes and to test their understanding of Newton's third law, we used the FCI. This is a set of 29 qualitative multiple choice questions on mechanics. Our personal experience with individual students is consistent with Hestenes's claim that success in this test correlates with a good understanding of Newtonian mechanics as measured by detailed interviews. We gave the FCI both as a pre-test in the first week of the class and as a post-test in the last week of the class. This permits us to get a general overview of both the students' preparation in mechanics and the overall effect of the course.

A detailed study of FCI results nationwide by Hake[18] compares the performance of a large number of classes on the FCI. Hake's results show an interesting uniformity. When the class's gain on the FCI (post-test average - pre-test average) is plotted against the class's pre-test score, classes of similar structure lie approximately along a straight line passing through the point (100,0). This is shown schematically in Fig. 2.

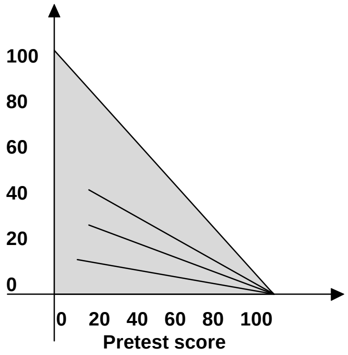


Fig. 2: Schematic of the Hake plot. A class's average pre-test FCI score is plotted as the abscissa, the pre- to post-test gain as the ordinate. Since the maximum average is 100%, every data point must lie in the shaded region. The lines of constant *h* are shown for a few values of *h*.

Traditional classes lie on the line closest to the horizontal axis and show limited improvement. The middle line represents classes with active engagement. The steepest line represents classes with active engagement and a research-based text. The negative slope of the line from a data point to the point (100,0) is a figure of merit:

$$h = (\text{class post-test average - class pre-test average})/(100 - \text{class pre-test average})$$

The interpretation of this is that two classes having the same figure of merit, *h*, have achieved the same *fraction of the possible gain*.

The FCI contains a well-defined set of four questions that deal with Newton's third law. (They are given in the Appendix.) In order to evaluate the overall effect of the course, we calculate the figure of merit for the class's average on this set of questions. Since at present no data has been presented to suggest that a result similar to Hake's might hold for a sub-cluster of the FCI, we present both the absolute results and the fraction of the possible gain.

Table 1: Lecture classes tested

Section	***N[1]***	***Class structure***	***FCI pre***	***FCI post***	***h***
A1	100	recitation[2]			
A2	38	tutorial (no MBL)[3]	47.8	66.7	0.36
A3	**109**	**tutorial**	**54.5**	**72.8**	**0.40**
B1	27	recitation	51.2	65.5	0.29
B2	19	recitation	58.8	69.1	0.25
C1	35	recitation	41.8	54.2	0.21
C2	18	recitation	38.3	47.5	0.15
D1	**69**	**tutorial**	**50.3**	**67.5**	**0.35**
D2	**48**	**tutorial**	**44.4**	**61.9**	**0.31**
E	42	recitation	55.4	55.9	0.01
F	**55**	**tutorial**	**53.9**	**67.8**	**0.30**

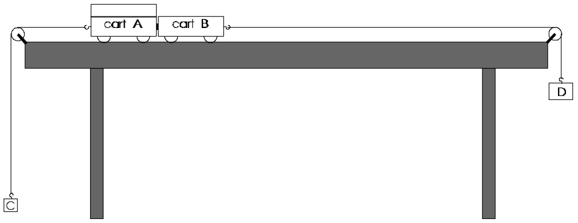
[1] N is the total number of students that took both the pre- and post tests except in the case A1 where it represents the total number of students in the class.

[2] In this class, only the VQ were given. The FCI was not done.

[3] In this class, tutorials were done, but the velocity and Newton-3 MBL tutorials were not given.

Two lecture classes taught by the same professor were tested with the VQ. Ten lecture classes run by six lecturers were tested with the FCI. The situations are summarized in Table 1. Each of the letters A to F specifies a particular instructor. When an instructor participated in the study in more than one semester, a number is assigned as well to allow us to specify a particular data set. Here and in the remainder of the tables, the rows corresponding to classes that received instruction using MBL are highlighted using bold.

The long-answer final exam question was constructed to require students to both construct a velocity graph describing a complex situation and to use Newton's third law appropriately. The question is shown in Figure 3. It was administered to one traditional lecture section (B2: N=50) and one tutorial lecture class (D2: N=82).[19] Although the content covered in this question clearly matches what all the students had been taught, it is both challenging and different than what was seen during the semester. Note that students were asked to explain their reasoning as part of their answers.



Two carts, A and B ($\text{Mass}_A > \text{Mass}_B$), are placed on a table then stuck together with Velcro. Using pulleys, two small blocks, C and D ($\text{mass}_C < \text{mass}_D$), are connected by light strings to the carts as shown below. Initially, the carts are held in place. **Ignore all friction in this problem.** At t=0, the carts are released. At t=3 seconds, the Velcro pulls apart and the 2 carts separate. At some later time, cart A returns to its starting point.

a. Draw and label 2 separate free-body diagrams, one for each cart, for a time after the carts start moving but before the Velcro pulls apart.

b. Rank all the horizontal forces from both your diagrams by magnitude. Explain the reasoning that you used.

c. Briefly describe the motion of cart A from t=0 until it returns to its starting point. On the graph provided, qualitatively sketch the velocity vs. time for this time period.

Fig 3: Long-problem exam question requiring both construction of a velocity graph and application of Newton's third law.

V. RESULTS

In this section we describe the results obtained. We first discuss the results of the multiple choice questions, beginning with a presentation of the overall FCI results to provide a normalization of the overall effectiveness of the tutorial environment for general concept building. We then present the specific results of the VQ and of the Newton 3 cluster of the FCI. Finally, we discuss the implications of the free-response-problem results. Note that all evaluations were not used in all classes.

A. Multiple Choice

Overall FCI

We display the results of pre- and post-FCI tests in tutorial and non-tutorial classes in Table 1. Ten of our classes gave the FCI as pre- and post-tests. Five were taught with tutorials, five with recitations. The data shown are *matched*, that is, only those students who took both the pre- and post-tests are included. The number of matched students is listed in Table 1 under N. Comparing the averages of the pre-test scores for all students taking the pre-test with the matched subset show that there is not a significant selection.

The results are displayed as a figure of merit (*h*) histogram in Fig. 4. The classes taught without tutorials are shown as solid bars, while those taught with tutorials are shown as gray bars.

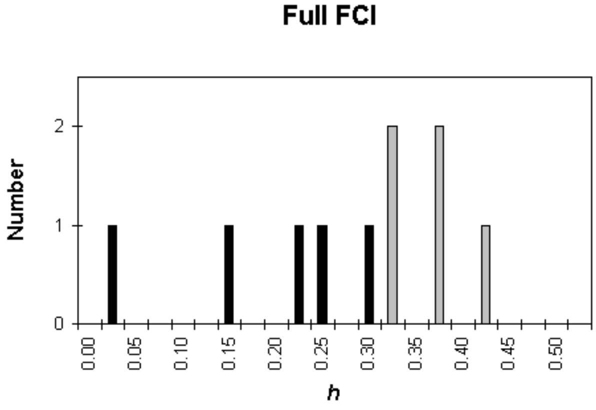


Fig. 4: Figure of merit histogram. *h* = fraction of possible gain obtained on the full FCI, for tutorial (gray bars) and non-tutorial (solid bars) lecture classes.

The tutorial classes systematically produced better overall FCI gains than the non-tutorial classes. The average fractional gains of the classes are

$<h> = 0.18$ (5 classes, with recitations)
$<h> = 0.35$ (5 classes, with tutorials)

Note that this average is taken as equally weighted over lecture classes, not by student. Every tutorial class had a larger h than every non-tutorial class.[20] Even the tutorial results are somewhat disappointing, achieving only about 1/3 of the possible gain. Both results, however, are consistent with Hake's survey (ref. 18). The non-tutorial scores are consistent with those of other traditional classes, and the tutorial results are about halfway between those of traditional and highly interactive non-traditional classes.

Assuming that all 10 classes are drawn from the same population, the probability that the shift of the means is random is less than 2% using a 2-tailed t-test with pooled variance.[21] If class E is excluded as an outlier, the probability that the shift in the means is random is less than 1%.

The same amount of instruction was offered students in both environments (3 hours of lecture and 1 hour of small class section). The primary difference in the tutorial and traditional classes is that the tutorial classes spend one hour per week on explicit concept building in a small-class group-learning-style environment, while the traditional classes spend one problem solving hour per week in a small-class lecture-style environment.

Velocity

The VQ were given in two of our classes taught by the same professor. In class A1, the professor (an award winning teacher and a popular lecturer) did his best to teach the material explicitly in lecture, devoting nearly three full lecture hours to the topic of instantaneous velocity. Lecture demonstrations with the same MBL apparatus as in the tutorial were used in a careful demonstration with much student interaction and discussion. The professor had the students watch and plot the professor's motion as he walked a variety of paths, and a number of problems relating to students' personal experience were presented, but no worksheets were distributed. In recitation sections, graduate teaching assistants spent one hour going over textbook problems on the same material.

Table 2: Percentage error on the VQ with and without MBL.

Instruction without MBL	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*
UMd[1]	59	7	62	37	14
Tufts[2]	18	7	41	18.5	17
6 school av.[3]	41	17	63	37	6
Instruction with MBL	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*
UMd[4]	**16**	**2**	**30**	**19**	**5**
Tufts[2]	**2**	**--**	**7**	**5**	**3**
6 school av.[3]	**11.5**	**2**	**13**	**11**	**7**

[1] UMd, prof. A, no tutorial (N = 100)
[2] Thornton and Sokoloff (N = 177 reported in ref 3)
[3] Thornton, (N = 505 reported in ref 5)
[4] UMd, prof. A, MBL tutorial, (N = 161)

In class A3, the tutorial system was in place, and one hour of tutorial was given as described in section III. The professor reduced the lecture time on the topic to a single hour, which was more typical of a traditional lecture and had little student interaction.

In both classes, the questions were given as part of an examination and, contrary to Thornton and Sokoloff, were not previously given to the students as homework. The results for the error rates are given in Table 2 and shown in Fig. 5.

The Maryland results with four hours of traditional instruction and no tutorial (class A1) resembled the 6-school average of traditional lecture classes reported in Thornton's lecture at the Raleigh conference (ref. 5). The Maryland result with one hour of MBL tutorial and one hour of lecture was substantially improved, but not as good as the improvement shown with four hours of MBL.

These results are consistent with those given by Thornton and Sokoloff. The fact that these results have been obtained with both the lecturer and the time of instruction controlled strongly supports the results in Thornton and Sokoloff and answers our question Q1 in favor of the MBL. The MBL activities play a significant role in the improvement of student understanding of the concept of velocity. It is not simply the extra time that is responsible. It also suggests as a partial answer to our question Q3: simply enhancing lectures is not effective in producing an improvement in the learning of the velocity concept for a significant number of students.

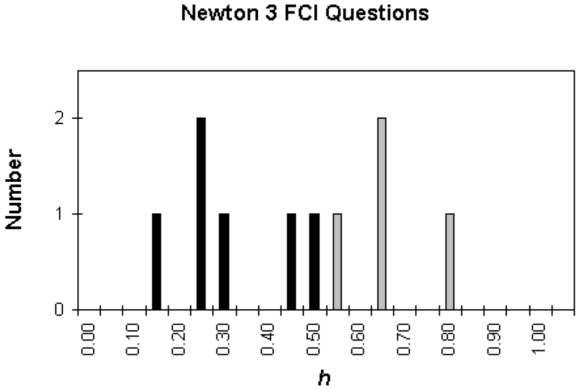


Fig. 6: Histogram of average figures of merit for the Newton 3 FCI cluster. Solid bars are for classes not using the MBL tutorial, gray bars for those using the MBL tutorial.

Newton 3

The Newton 3 tutorial was evaluated using the four FCI questions 2, 11, 13, and 14 (N3 FCI). The results are given in Table 3 and shown as a

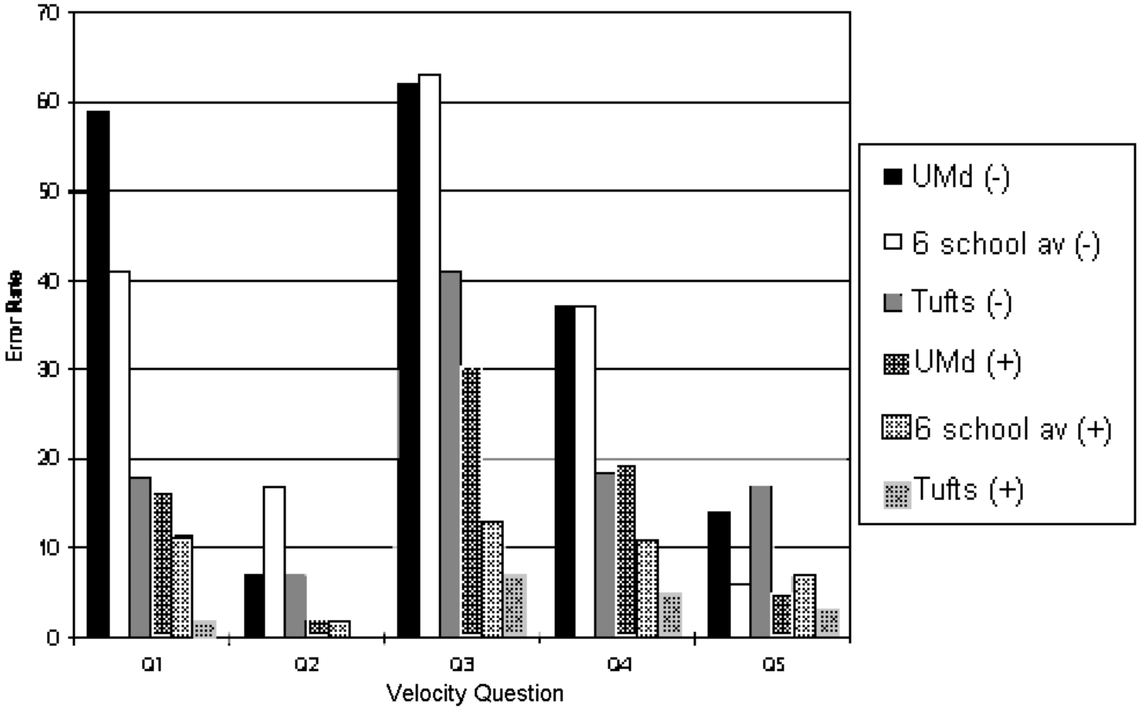


Fig. 5: Error rate on the velocity questions (VQ).

histogram in Fig. 6. The table gives the fraction of (matched) students answering each of the N3 FCI questions correctly at the beginning (pre) and end (post) of the semester. A figure of merit, *h* = (class post-test average- class pre-test average)/(100 - class pre-test average), is calculated *for each question* in analogy with the Hake figure of merit for the full FCI. The four *h*-values are then averaged in the last column to give a figure of merit for the Newton 3 cluster, h_{N3}.

The results are systematically better for the tutorial classes. Indeed, every tutorial class has a higher value of h_{N3} than every non-tutorial class (though a similar statement is not true for the *h*-values for every individual question). The average values of h_{N3} for each cluster of classes are:

$\langle h_{N3} \rangle = 0.28$ (6 classes, no N3 MBL tutorial)
$\langle h_{N3} \rangle = 0.64$ (4 classes, with N3 MBL tutorial)

In the first semester in which tutorials were tested, there was no tutorial specifically oriented towards Newton 3. Our tutorial was written for the subsequent semester. As a result, the first Maryland tutorial class, A2, used tutorials but not a Newton 3 MBL tutorial. This, therefore, gives us a control for individual lecturer as well as for the presence of tutorials. (No special effort was devoted to Newton 3 in lecture in either case.) The result is:

$\langle h_{N3} \rangle = 0.40$ (A2: no N3 MBL tutorial)
$\langle h_{N3} \rangle = 0.65$ (A3: with N3 MBL tutorial)

B. Long Problem

The long-exam problem shown in Fig. 3 was given in one tutorial class (D2) and one non-tutorial class (B2). Overall, performance on the problem was better for the tutorial than for the non-tutorial students. However, in this paper we will only discuss issues related to the velocity graph and Newton's third law.

Velocity

Part of the examination question asked the student to generate a velocity vs. time graph for a complicated situation. The critical elements of a fully correct solution show the velocity starting at 0, increasing linearly until t=3 seconds, and then decreasing linearly to some negative value.[22]

Students from both classes struggled with this question. Table 4 shows a breakdown of student responses. Only a small fraction of the students in either class were able to draw a graph that reflected the critical features, but the tutorial students did better than the students in the recitations. After traditional instruction, 12% of the students drew a correct graph. After MBL tutorials, 22% of the students drew a correct graph.

Table 4: Results on student construction of velocity graph

Class	***2***			***11***			***13***			***14***			h_{N3}
	Pre	**Post**	***h***	**Pre**	**Post**	***h***	**Pre**	**Post**	***h***	**Pre**	**Post**	***h***	
A2	26%	66%	0.54	53%	82%	0.61	37%	34%	-0.04	63%	82%	0.50	0.40
A3	**39%**	**89%**	**0.82**	**44%**	**87%**	**0.77**	**14%**	**52%**	**0.44**	**73%**	**88%**	**0.57**	**0.65**
B1	22%	67%	0.57	41%	78%	0.62	18%	52%	0.41	63%	74%	0.30	0.48
B2	37%	53%	0.25	47%	68%	0.40	26%	58%	0.43	68%	68%	0.00	0.27
C1	23%	32%	0.12	31%	46%	0.21	21%	43%	0.28	54%	66%	0.25	0.22
C2	39%	44%	0.09	39%	39%	0.00	28%	28%	0.00	50%	67%	0.33	0.11
D1	**38%**	**91%**	**0.86**	**45%**	**90%**	**0.82**	**29%**	**70%**	**0.57**	**59%**	**93%**	**0.82**	**0.77**
D2	**35%**	**66%**	**0.47**	**40%**	**73%**	**0.55**	**21%**	**46%**	**0.32**	**58%**	**90%**	**0.75**	**0.52**
E	19%	45%	0.32	52%	60%	0.15	24%	40%	0.22	60%	64%	0.12	0.20
F	**26%**	**76%**	**0.68**	**46%**	**85%**	**0.73**	**24%**	**56%**	**0.43**	**60%**	**89%**	**0.73**	**0.64**

Table 3: Results on the FCI Newton 3 questions. Given for each class are the percentage of students answering each question at the beginning (pre) and end (post) of the class. For each question, h is calculated to be (post-pre)/(100-pre). The column headed h_{N3} gives the average of the four h values in each row. Classes using the N3 MBL tutorial are indicated in bold. (Note class A2 used tutorials, but not the MBL ones.)

in long exam question.

	% correct	% apparently correct, but ending at $v=0^{b)}$	% other incorrect response
D2 (N=50)	12	10	78
B2 (N=82)	**22**	**21**	**57**

a) These students demonstrated understanding of the critical features of the graph.
b) While showing some of the critical features of a correct graph, these students mistakenly ended the graph at v=0, often citing the return of the cart to its initial position as the reason.

Analysis of the incorrect graphs along with the accompanying explanations revealed some of the students' difficulties. Many students showed in a variety of ways that they had the well-documented confusion between position and velocity. Some drew graphs that at first glance appear correct: the graph increased linearly for the first 3 seconds and then decreased linearly after. However the graph ended at v=0, and some of these students indicated that this coincided with the cart returning to its starting location. Many drew graphs that had incorrect combinations of linear segments, including discontinuities in velocity. Others drew dramatically curved features in their velocity-time graphs. Most of these graphs indicated severe conceptual difficulties even if interpreted as a position vs. time graph. It is worth noting that it is clear from many of their explanations that the students intended to draw a velocity vs. time graph.

Both the percentage of correctly drawn graphs and the nature of the incorrect graphs confirm that while student difficulties understanding kinematics is pervasive even after instruction, the modified instruction described earlier in this paper appears to be helping address these difficulties. Although the VQ were not given in these classes, approximately 70% of the students in the comparable tutorial class A3 answered all of the multiple choice questions correctly, while only about 40% of those in the recitation class A1 answered them all correctly. The relative results on the long-problem are qualitatively consistent with the results of the VQ, but the absolute number of students getting correct answers on the long-problem was substantially lower (22% of the tutorial students correct vs. 12% of recitation students correct). Since no classes were evaluated with both the VQ and the long problem, we cannot completely answer Question Q2, but our indications are that the VQ may not suffice. Our results suggest that answering multiple-choice questions correctly is not sufficient to guarantee a robust and fully functional understanding of the relevant concepts for a significant number of students.

Newton 3

Another part of the same examination question tested student facility with dynamical concepts. The students were asked to draw a free body diagram of each cart shown in Fig. 3 and to rank the magnitudes of the horizontal forces. Note in particular that by Newton's third law, the magnitude of the force of cart A on cart B is equal to that of cart B on cart A.

The breakdown of student responses to this part of the question is shown in Table 5. In the tutorial classes, 55% of the students correctly identified and compared the third law force pair. In the non-tutorial class 42% identified and correctly compared these forces.[23] (This result favoring the tutorial class is particularly notable since their pre-test N3 FCI scores were lower than the recitation classes's score, 38% correct to 44% correct.) Many students identified that the two carts were exerting forces on one another, but stated explicitly that the two forces were not of equal magnitude. In addition, there were also many students who did not even recognize that the two carts exert forces on each other. This was particularly common in the non-tutorial class.

These results should be compared with the results on the post-test N3 FCI questions for the same two classes, 69% and 62% respectively. The discrepancy between the multiple-choice and long-answer problems (in this case both questions were done by both groups) also suggests that the answer to question Q2 might be: the short answer results provide an indication, but overestimate the students' knowledge.

Table 5: Results on student use of Newton's third law in long exam question.

	% correct	% used the same symbol but did not compare forces	% stated third law force pair have different magnitudes	% no identification of contact forces	% other incorrect response
D2 (N=50)	42	6	22	14	16
B2 (N=82)	**55**	**0**	**40**	**1**	**4**

VI. SUMMARY AND CONCLUSIONS

In this paper we have discussed an experiment to test the effectiveness of replacing one hour of problem-solving recitation by one hour of active-engagement MBL addressing the issues of instantaneous velocity and Newton's third law delivered in a University of Washington style tutorial.

The velocity issue was probed by using the multiple choice velocity graph-matching questions given in ref. 3 in two classes taught by the same professor. In one class, the material was taught in lecture with additional lectures given on the subject and the professor doing his best to "teach to the test" without actually doing the test questions in class. In a second class, the professor ignored the test but a single hour of tutorial based on *Tools for Scientific Thinking* was given. In the non-tutorial class the results were very close to the six school average of lecture classes reported in ref. 5. In the tutorial class, the error rates fell by more than a factor of 2 for all questions. Although this result is not as dramatic as those produced by Thornton and Sokoloff after four additional hours of MBL laboratory, the results are still impressive, especially since we controlled for both the instructor and the time of instruction.

In our second experiment, we constructed a tutorial using MBL on the subject of Newton's third law. In this case, we used the four relevant questions from the FCI in pre- and post-testing as an evaluation tool. Of the ten classes tested, the tutorial was given in 4 lecture classes with three different professors, and it was not given in 6 lecture classes with three different professors. One of the professors taught a class in each group giving us a specific control for instructor. Both the absolute gains and the final total scores favor the tutorial classes with every tutorial class scoring a higher fraction of the possible gain (h_{N3}) than every non-tutorial class. The professor who did both a recitation and tutorial section, found his class's value of h_{N3} increase by 60% when he used the tutorial.

We therefore conclude that our answer to question Q1 is: targeted MBL tutorials can be effective in helping students build conceptual understanding, but do not provide a complete solution to the problem of building a robust and functional knowledge for many students.

A long problem requiring the application of the velocity concept, the building of a velocity graph, and the application of Newton's third law in a complex situation was also given to one tutorial and one recitation class. The tutorial students performed better than the recitation students. In the N3 case where both short and long answer data were available for the same class, the long answer results favored the tutorial students slightly more strongly than the multiple choice questions. But in all cases, the number of students able to produce the correct concept in a complex situation was significantly less than suggested by the multiple choice questions. This indicates that the answer to question Q2 is: multiple choice tests are qualitatively indicative of the direction of change, but cannot be used to determine the extent of robust and functional knowledge developed by the class.

In this experiment we did not test for "side effects". Since the MBL activities were added at the expense of problem-solving recitations we should also test whether there was a deterioration in problem-solving for students who did tutorials instead of recitations. We do not expect a significant effect as our personal anecdotal evidence suggests that recitations are effective for only a small fraction of students. This should, however, be tested in more detailed studies. There are strong indications from earlier work[24] that successful problem solving at the introductory level is often not associated with a growth in conceptual understanding. It may be that only a small fraction of students can successfully learn physics in the order: algorithms first, understanding second; and that it would be more effective for most students to reverse the order.

Thornton and Sokoloff conjectured that the MBL activities they had designed were unusually effective for five reasons:

1. Students focus on the physical world.
2. Immediate feedback is available.
3. Collaboration is encouraged.
4. Powerful tools reduce unnecessary drudgery.
5. Students understand the specific and familiar before moving to the more general and abstract.

These conjectures are consistent with modern theories of learning,[25] including those built on the work of Piaget and Vygotsky, and on our current understanding of the structure of short and long-term memory buffers.[26] To this list we add a sixth conjecture:

6. Students are actively engaged in exploring and constructing their own understanding.

The Thornton-Sokoloff conjectures appear to be confirmed by a variety of anecdotes describing the success of the substitution of active-engagement MBL activities for traditional labs, and by the failure of the same equipment when used as traditional labs without the engagement/discovery component.[27] These have not, unfortunately, been documented in the literature. It would be useful to have additional detailed experiments comparing different methods in order to build an understanding of exactly what components of MBL activities are proving effective.

Since we relied on all of the Thornton and Sokoloff conjectures (and one of our own) in building our units, we are unable to distinguish which of the elements are critical.[28] Note that the impact of all of the six conjectures could well be achieved without the use of MBL equipment. It would be most interesting to carry out additional large scale studies of the effectiveness of active engagement activities that do not include MBL on the learning of specific concepts.

ACKNOWLEDGMENTS

One of us (EFR) would like to thank Cliff Swartz for a number of thought-provoking discussions probing the Thornton and Sokoloff results. We would also like to thank Ron Thornton, David Sokoloff, and Michael Wittmann for useful discussions and comments on the paper.

Work supported in part by NSF Grants RED-9355849 and DUE-9455561 Computer facilities provided by NSF ILI Grant DUE-9550890 and by the University of Maryland Computer Science Center.

---

[1] Trowbridge, David E., and Lillian C. McDermott, "Investigation of student understanding of the concept of velocity in one dimension", *Am. J. Phys.* **48** (1980) 1020-1028.; Viennot, L., "Spontaneous reasoning in elementary dynamics", *Eur. J. Sci. Educ.*, **1** (1979) 205-221; Halloun, Ibrahim A., and David Hestenes, "Common sense concepts about motion", *Am. J. Phys.* **53** (1985) 1056-1065; Clement, J., "Students' Preconceptions in Introductory Mechanics", *Am. J. Phys.* **50** (1982) 66-71; Beichner, Robert, "Testing student interpretation of kinematic graphs," *Am. J. Phys.* **62** (1994) 750-762.

[2] In this paper, we use the term MBL to refer not simply to the use of a microcomputer to collect and display data, but to laboratory activities using microcomputers in which there is an attempt to actively engage students intellectually and to help them construct an understanding of the relevant concepts.

[3] Thornton, R. K., and D. R. Sokoloff, "Learning motion concepts using real-time microcomputer-based laboratory tools", *Am. J. Phys.* **58** (1990) 858-867.

[4] The poor results achieved by students on the velocity graph questions, despite traditional instruction in the subject, might be interpreted as difficulties reading graphs and not as a confusion on concept. However, the fact that students succeed at a much higher rate on position graphs, plus the reports of interviews from many researchers indicate that this is a confusion of concept, not just of representation.

[5] Thornton, R. K., "Tools for Scientific Thinking: Learning Physical Concepts with Real-Time Laboratory Measurement Tools", in *The Conference on Computers in Physics Instruction, Proceedings* E. F. Redish and J. Risley, Eds. (Addison-Wesley, 1990) 177-188.

[6] The textbook for all semesters reported was Tipler, Paul A., *Physics for Scientists and Engineers, 3rd Ed.* (Worth Publishers, NY, 1991).

[7] This was confirmed by the results of the Hestenes test reported below and by interviews associated with another project.

---

[8] L. C. McDermott, P. S. Shaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (University of Washington, Seattle, 1991-present). For a description of tutorials as used at the University of Washington and an experimental result using them, see McDermott, L. C., Peter S. Shaffer, and Mark D. Somers, "Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwoods's machine", *Am. J. Phys*. **62** (1994) 46-55; Shaffer, P. S., and L. C. McDermott, "Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of an instructional strategy", *Am. J. Phys.* **60** (1992) 1003-1013.

[9] Morse, Robert A., "The classic method of Mrs. Socrates", *The Physics Teacher* **32** (May, 1994) 276-277.

[10] McDermott, Lillian C., "Millikan Lecture 1990: What we teach and what is learned — Closing the gap", *Am. J. Phys.* **59** (1991) 301-315.

[11] We used Intel 386, 486, and Macintosh SE personal computers. The ULI, sonic ranger, and force probes are from Vernier software, Portland, OR. Only the Motion and Datalogger software that comes bundled with the ULI were needed. Two Pasco carts or their equivalent are also required for the Newton 3 tutorial. The current cost of the required equipment is about $2500 per station.

[12] Thornton, R. and D. Sokoloff, *Tools for Scientific Thinking* (Vernier Software, Portland OR, 1992 and 1993).

[13] Ref. 3, Fig. 2.

[14] Laws, P. R. Thornton, and D. Sokoloff, *RealTime Physics* (Vernier Software, Portland OR, 1995).

[15] Redish, Edward F., "Implications of cognitive studies for teaching physics", *Am. J. Phys.* **62** (1994) 796-803.

[16] Sandin, T. R., "On not choosing multiple choice", *Am. J. Phys.* **53** (1985) 299-300.

[17] Hestenes, D., M. Wells, and G. Swackhammer, "Force Concept Inventory", *The Physics Teacher* **30**:3 (1992) 141-158.

[18] Hake, R. R., "A five-thousand-student survey of mechanics test data for introductory physics courses", Indiana University preprint, April 1996, to be published.

[19] The numbers of students completing the final exam was greater than the number of students completing both the pre-and post FCI tests which were given in lecture.

[20] In absolute final scores, one tutorial class with a low pre-test score finished below some of the non-tutorial classes, and one non-tutorial class with a high pre-test score finished above some of the tutorial classes.

[21] Howell, David C., *Statistical Methods for Psychology, 3rd Ed.* (Duxbury Press, Belmont, California 1992) pp. 181-185.

[22] A graph reversed with respect to the horizontal axis would also be considered correct.

[23] A few students in the non-tutorial class used the same symbol for these two forces, but did not state whether the forces were equal, so it was impossible to determine if they were identifying these two forces as having equal magnitudes. Note that many students used the same symbol to represent forces which clearly had *different* magnitudes.

[24] Reif, Frederick, "Millikan Lecture 1994: Understanding and teaching important scientific thought processes" *Am. J. Phys*. **63** (1995) 17-32.

[25] See references in ref. 15.

[26] Schacter, Daniel L., "Memory", in *Foundations of Cognitive* Science, M. I. Posner, ed. (MIT Press, Cambridge MA, 1989) 683-725.

[27] Thornton, Ron, private communication, January 1994.

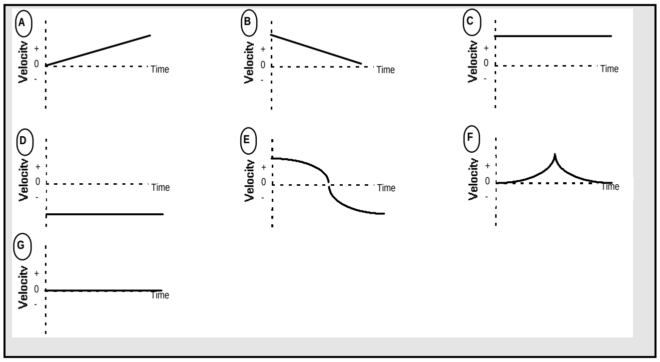
[28] From our personal teaching experience, we expect that number 6, active engagement, is the most significant.

Appendix: Questions Used in Evaluations

Thornton-Sokoloff Velocity graph question (VQ)

An object's motion is restricted to one dimension along the + distance axis. Answer each of the questions below by selecting the velocity graph that is the best choice to describe the answer. You may use a graph more than once or not at all.

a. Which velocity graph shows an object going away from the origin at a steady velocity?
b. Which velocity graph shows an object that is standing still?
c. Which velocity graph shows an object moving toward the origin at a steady velocity?
d. Which velocity graph shows an object changing direction?
e. Which velocity graph shows an object that is steadily increasing its speed?



FCI Newton 3 Questions

2. Imagine a head-on collision between a large truck and a small compact car. During the collision:

   (A) the truck exerts a greater amount of force on the car than the car exerts on the truck.

   (B) the car exerts a greater amount of force on the truck than the truck exerts on the car.

   (C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.

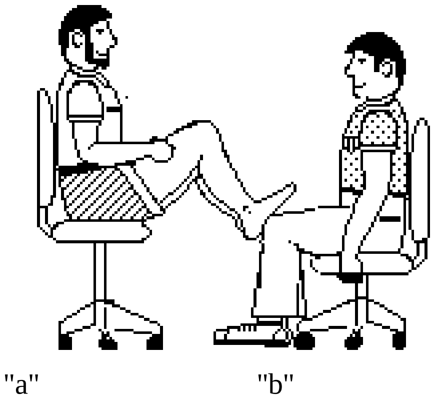
   (D) the truck exerts a force on the car but the car does not exert a force on the truck.

   (E) the truck exerts the same amount of force on the car as the car exerts on the truck.

11. In the figure at right, student "a" has a mass of 95 kg and student "b" has a mass of 77 kg. They sit in identical office chairs facing each other.

    Student "a" places his bare feet on the knees of student "b", as shown. Student "a" then suddenly pushes outward with his feet, causing both chairs to move.

    In this situation:



   (A) neither student exerts a force on the other.

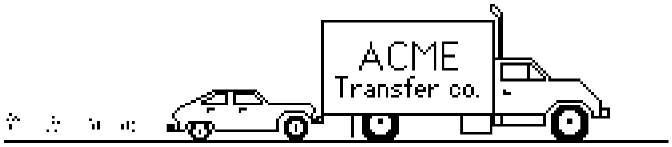
   (B) student "a" exerts a force on student "b", but "b" does not exert any force on "a".

(C) each student exerts a force on the other, but "b" exerts the larger force.

(D) each student exerts a force on the other, but "a" exerts the larger force.

(E) each student exerts the same amount of force on the other.

Refer to the following statement and diagram while answering the next two questions.



A large truck breaks down out on the road and receives a push back into town by a small compact car.

13. While the car, still pushing the truck, is **<u>speeding up</u>** to get up to cruising speed:

(A) the amount of force with which the car pushes on the truck is equal to that of the truck pushing back on the car.

(B) the amount of force of the car pushing on the truck is smaller than that of the truck pushing back on the car.

(C) the amount of force of he car pushing against the truck is greater than that of the truck pushing back on the car.

(D) the car's engine is running so it applies a force as it pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.

(E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

14. After the car reaches the constant cruising speed at which its driver wishes to push the truck:

(A) the amount of force of the car pushing on the truck is equal to that of the truck pushing back on the car.

(B) the amount of force of the car pushing on the truck is smaller than that of the truck pushing back on the car.

(C) the amount of force of the car pushing on the truck is greater than that of the truck pushing back on the car.

(D) the car's engine is running so it applies a force as it pushes against the truck, but the truck's engine is not running so it can't push back against the car. The truck is pushed forward simply because it is in the way of the car.

(E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.