An Analysis of Asynchronous Online Homework Discussions in Introductory Physics Courses

Gerd Kortemeyer*
Lyman-Briggs School of Science
and Division of Science and Mathematics Education
Michigan State University
East Lansing, MI 48825
(Dated: January 5, 2006)

Asynchronous online student discussions of online homework problems in introductory physics courses are analyzed with respect to course type, student course performance, student gender, problem difficulty, and problem type. It is found that these variables can significantly change the character of the online student collaborations.

PACS numbers: 01.40.Fk,01.40.Gm,01.50.Ht,01.50.Kw

I. INTRODUCTION

Students discussing physics with their peers in-class has proven to be an effective way of teaching [1], and the practice has found wide-spread acceptance. Using online forums, the practice can be extended outside the classroom. Over the past years, we have been using an online system where the threaded discussion forums are directly attached to randomizing online problems. Despite supporting research (e.g., [2] for a review), we continue to be surprised by the richness of the ensuing peer-interactions. In this study, we are attempting to systematically analyze the student discussion contributions, in particular with respect to properties of the courses, the students, and the problems. Our goal is to first identify online discussion behavioral patterns of successful students, and to then identify the problem properties which elicit them.

A. The LON-CAPA Online System

LON-CAPA started in 1992 as a system to give randomized homework to students in introductory physics courses. “Randomized” means that each student sees a different version of the same computer-generated problem: different numbers, choices, graphs, images, simulation parameters, etc, see Fig. 1. Randomization was implemented as a means to both control “cheating” and foster student collaboration on a conceptual level — since problem answers will differ from student to student, learners cannot simply exchange the correct answers when collaborating with each other.

LON-CAPA allows for immediate feedback on problem correctness to the student, as well as multiple tries to arrive at the correct solution (both features could be disabled by the instructor). The system is designed to foster communication among the learners, and asynchronous threaded discussion boards are attached directly to the bottom of every online resource. For the purposes of this project, it is therefore possible to establish a one-to-one association between an online problems and discussions.

Students can post anonymously or using a screenname, however, the full name is always visible to the instructors (students know this). Also, occasionally, instructors post to the discussion. Over time, competing discussion sites developed outside of LON-CAPA, which are completely anonymous and are not visited by instructors. Kashy [3] found that the use of the internal discussion sites is positively correlated to course grades and FCI scores, while the use of the external sites is negatively correlated to these scores.

In addition, LON-CAPA keeps statistical data for every problem, which allows instructors to associate problems with their degree of difficulty.

B. Courses

Discussions from three courses at Michigan State University were analyzed, namely, the first semester of an algebra-based course with students from a wide variety of majors, as well as the first and the second semester of a calculus-based course with a majority of pre-medical students. In both courses, the complete teaching material was provided online, with homework problems embedded. No textbook was required in either course. The algebra-based course had one section that was completely taught online, but the majority of the students in the algebra-based course, and all students in the calculus-based course, had regular lectures throughout the week. In the case of the calculus-based course, a parallel lab was offered. All three courses were graded on an absolute scale without “curving,” and student collaboration was explicitly encouraged. Homework contributed to less than 20 percent to the final grade.

A total of 134 online problems with 1367 associated discussion contributions were analyzed in the first

*Electronic address: `korte@lite.msu.edu`; URL: `http://www.lite.msu.edu/kortemeyer/`

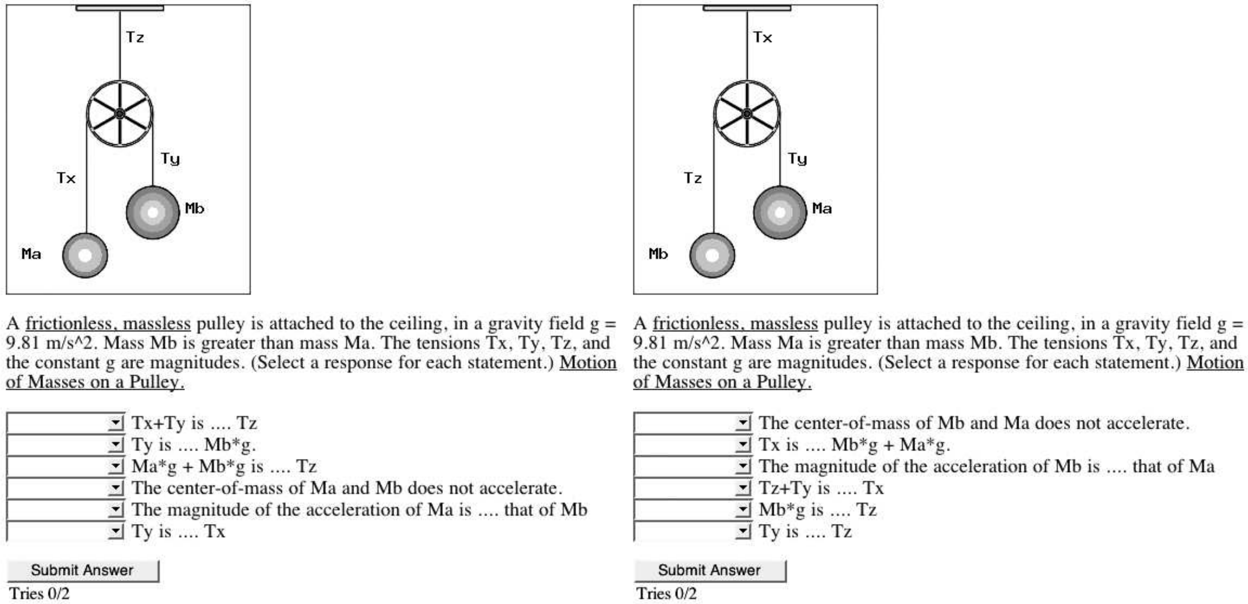


FIG. 1: Web-rendering of the same LON-CAPA problem for two different students.

semester of the algebra-based course, as well as 215 problems with 1078 discussion contributions in the first semester, and 148 problems with 949 discussion contributions in the second semester of the calculus-based course.

In addition, within the first semester calculus-based course (enrollment: 211 students (82 men, 129 women)), discussion characteristics were correlated to student characteristics.

II. METHODOLOGY

A. Problem Classification

Kashy [4] showed that student mastery of different types of homework problems correlates differently with the students' performance on final exams — with multiple-choice non-numerical problems having the lowest correlation, and numerical/mathematical problems that require a translation of representation having the highest. Steinberg [5] also analyzed student performance on multiple-choice diagnostics and open-ended exam problems, and found that while those correlate in general, for certain students and certain problems, responses differ greatly. For this project, we chose a finer-grained classification scheme of problem types: Redish [6] identifies eight classes and features of exam and homework problems, and an adapted version of this scheme will be used:

Multiple-choice and short-answer problems The most basic and most easily computer-evaluated type of problem, representing the conventional (typical back-of-chapter textbook) problem.

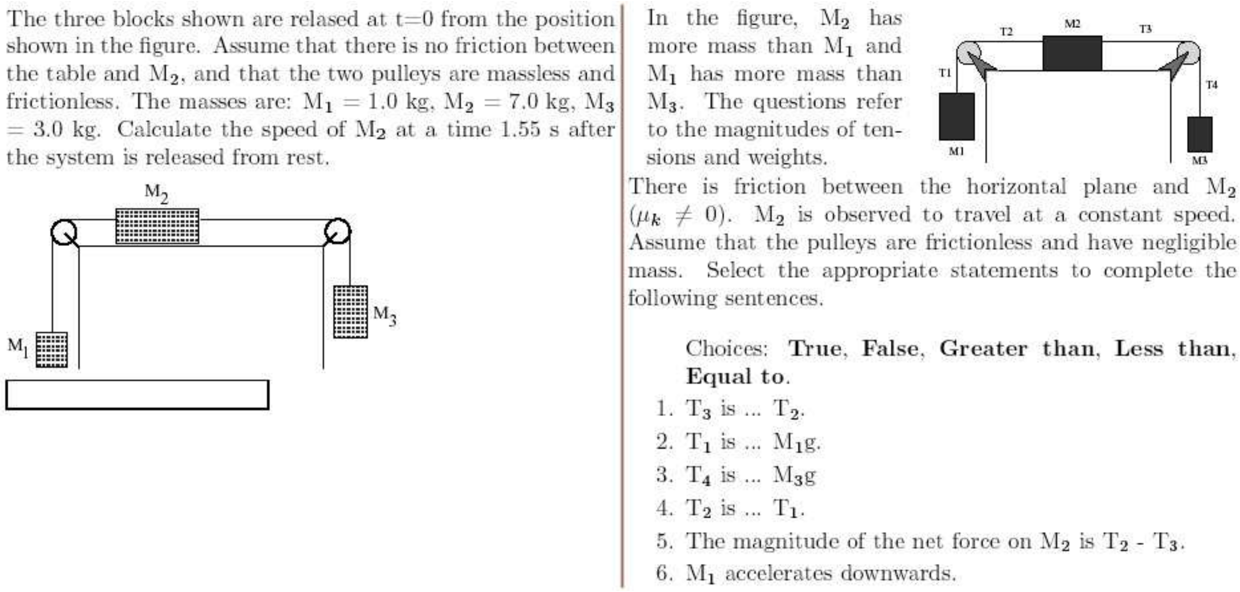
For the purposes of this project, "multiple choice" and "short-answer" will be considered as separate classes, where short-answer includes numerical answers such as "`17 kg/m^3`," and formula answers, such as "`1/2*m*(vx^2+vy^2)`." The problems on the left side of Figs. 2 and 3 are examples of "short-(numerical)-answer" problems.

Multiple-choice multiple-response problems This type of problem, a first step beyond conventional problems, requires a student to evaluate each statement and make a decision about it. The problem on the right side of Fig. 2 is of this type.

Representation-translation problems This type of problem requires a student to translate between different representations of the same situation, for example from a graphical to a numerical or textual representation. The answer might be required in different formats, for example in the problem on the right side of Fig. 3, it is a short-numerical-answer. Translation between representations can be surprisingly challenging for physics learners [7, 8].

For the purposes of this project, "representation-translation" will be considered a feature, which may or may not apply to any of the other problem types.

Ranking-tasks This type of problem requires a student to rank a number of statements, scenarios, or objects with respect to a certain feature. For example, a student might be asked to rank a number of projectiles in the order that they will hit the ground, or a number of locations in order of the strength of their local electric potential.



The three blocks shown are relased at t=0 from the position shown in the figure. Assume that there is no friction between the table and M_2, and that the two pulleys are massless and frictionless. The masses are: $M_1 = 1.0$ kg, $M_2 = 7.0$ kg, $M_3 = 3.0$ kg. Calculate the speed of M_2 at a time 1.55 s after the system is released from rest.

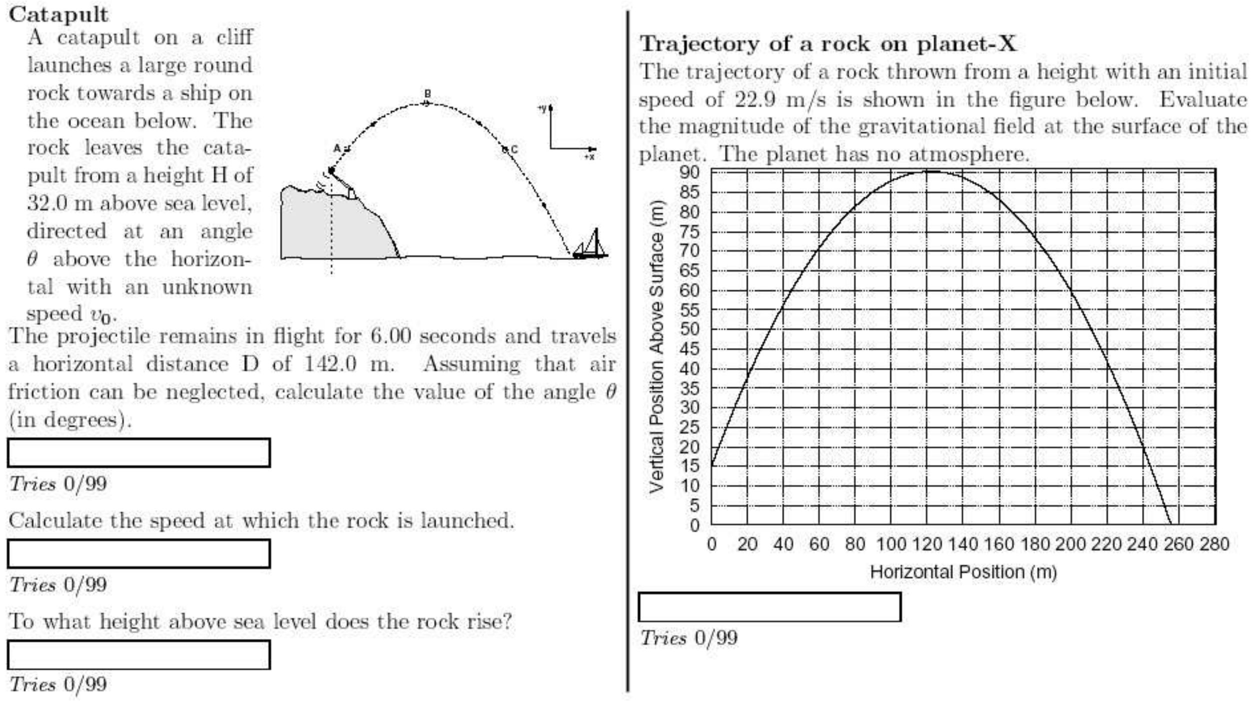
In the figure, M_2 has more mass than M_1 and M_1 has more mass than M_3. The questions refer to the magnitudes of tensions and weights.

There is friction between the horizontal plane and M_2 ($\mu_k \neq 0$). M_2 is observed to travel at a constant speed. Assume that the pulleys are frictionless and have negligible mass. Select the appropriate statements to complete the following sentences.

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

1. T_3 is ... T_2.
2. T_1 is ... M_1g.
3. T_4 is ... M_3g
4. T_2 is ... T_1.
5. The magnitude of the net force on M_2 is T_2 - T_3.
6. M_1 accelerates downwards.

FIG. 2: Example of two LON-CAPA problems addressing the same concepts. The problem on the left is a conventional short-numerical-answer problem, while the problem on the right is of type "multiple-choice multiple-response."



Catapult

A catapult on a cliff launches a large round rock towards a ship on the ocean below. The rock leaves the catapult from a height H of 32.0 m above sea level, directed at an angle θ above the horizontal with an unknown speed v_0.

The projectile remains in flight for 6.00 seconds and travels a horizontal distance D of 142.0 m. Assuming that air friction can be neglected, calculate the value of the angle θ (in degrees).

Tries 0/99

Calculate the speed at which the rock is launched.

Tries 0/99

To what height above sea level does the rock rise?

Tries 0/99

Trajectory of a rock on planet-X

The trajectory of a rock thrown from a height with an initial speed of 22.9 m/s is shown in the figure below. Evaluate the magnitude of the gravitational field at the surface of the planet. The planet has no atmosphere.

Tries 0/99

FIG. 3: Example of two LON-CAPA problems addressing the same concepts in two different representations. The problem on the left is a conventional short-numerical-answer problem, while the problem on the right requires "representation-translation."

Context-based reasoning problems The distinguishing characteristic of these problems is that they are set in the context of real-world scenarios and not in the context of the artificial "zero-friction" laboratory scenarios of typical textbook problems.

As in the case of "representation-translation," "context-based-reasoning" in this project will be considered a feature, which may or may not apply to any of the other problem types.

Estimation problems , also known as "Fermi Problems," require the student to form a model for a scenario, and make reasonable assumptions. A typical example is "How many barbers are there in

Chicago?" or "How long will I have to wait to find a parking spot?" Students do need to explain their reasoning.

While students find it initially hard to believe that these problems have anything to do with physics, hardly any expert physicist would deny their significance in learning how to solve problems [9].

Qualitative problems This type of problem asks students to make judgments about physical scenarios, and in that respect are somewhat similar to ranking problems. While the problems themselves are of the type "Is this high enough?" or "Can we safely ignore . . . ?," they often do require at least "back-of-the-envelope" calculations to give informed answers. As in the case of estimation problems, students have to explain their reasoning, but the problem itself is usually more structured, and at least the initial answer is more easily evaluated by a computer.

Essay problems These are "explain why" problems. A certain scenario is presented, and students are asked to explain why it turns out the way it does. Students are not asked to recall a certain law — it is given to them. Instead, they are asked to discuss its validity.

All 497 online problems available for this study were classified by the author. The three courses did not include estimation, qualitative, and essay problems, which cannot be graded automatically within the online system. Table I shows the classification distribution of the online problems available for this project.

None of the problems required context-based reasoning or expected a free-form short textual answer. Approximately 14 percent of the problems required representation translation. The vast majority of problems were conventional numerical problems, which expect a numerical answer with associated physical unit.

The difficulty index for each problem was computed according to the formula

$$\text{Difficulty Index} = 10\left(1 - \frac{N_{\text{correct}}}{N_{\text{attempts}}}\right)$$

where N_{correct} is the total number of correct solutions of the problem in the course, and N_{attempts} is the total number of correct and incorrect solution submissions (the system allows multiple attempts to arrive at the correct solution, see subsection I A). If all submissions were correct, meaning, every student would have solved the problem correctly on the first attempt, the difficulty index would be 0. If none of the submissions were correct, the index would be 10.

B. Discussion Classification

To perform a quantitative discourse analysis of the online discussions, the student discussion entries were classified into three types and four features. The four types are

Emotional - discussion contributions were classified as "emotional" if they mostly communicated opinions, complaints, gratitude, feelings, etc. Two subtypes were "positive" and "negative."

Surface - discussion contributions were classified as "surface" if they dealt with surface features of the problem or were surface level requests for help. Two subtypes were "question" and "answer."

Procedural - contributions that describe or inquire about a mechanisms to solve the problem without mention of the underlying concepts or reasoning. Two subtypes were "question" and "answer."

Conceptual - contributions that deal with the underlying concepts of the problem. Two subtypes were "question" and "answer."

In addition, discussion contributions were classified by the following features:

Unrelated - the contribution is not related to the problem.

Solution-oriented - the goal of the contribution is to arrive at the correct answer without mentioning or dealing with the mathematics or physics of the problem.

Mathematical - the contribution deals mostly with the mathematical aspects of the problem.

Physics - the contribution deals mostly with the physics aspects of the problem.

Table II shows examples of contributions and their classification. Each combination of subtype and feature forms a "class" in the analysis.

This coding scheme has to the author's knowledge not been previously used in literature, but was chosen in correspondence to the observations reported in [10–12] to distinguish between desirable and undesirable problem solving strategies. Clearly, instructors would want their students to work on a conceptual physics level, yet oftentimes students categorize problems according to surface features [11], and attempt to proceed in a purely procedural approach ("plug-and-chug") to as quickly as possible arrive at the correct solution [10]. Pascarella [12] reports that online homework tends to affirm students in this undesirable approach. All 3394 discussion contributions were classified by the author over the course of two months. Discussion contributions were always classified as a whole, and since they were mostly fairly short

TABLE I: Classification of the online problems according the classification scheme described in subsection II A (adapted from Redish [6]). The columns denote the different problem types, while the rows denote the features of required representation translation and context-based reasoning.

	Multiple-choice and short-answer				Mult.-choice mult.-resp.	Ranking	Click-on-image	
	Multiple-choice	Textual	Numerical	Formula				
"Conventional"	10		355	3	54	4	2	428
Rep-Trans	7		38		16	1	7	69
Context-based								0
	17		393	3	70	5	9	497

TABLE II: Examples of discussion contribution types and features.

	Unrelated	Solution	Math	Physics
Emotional	Why is it that homeworks are getting longer and longer?	Everyone keeps saying they figured it out, but no one is telling how. Please let us know because we have tried everything!	Bless your heart, and thank you for having the patience to explain this vector addition stuff to people like me who're really struggling with this vector and sin, cos stuff. It's starting to all come together.	Sometimes, when I think of the word "physics," I get a sickening feeling in the pit of my stomach. It's sort of like a burning sensation.
Surface	If this is extra credit, does that mean it won't be on the exam?	Post the answers you know are correct for sure ... all do this .. and we'll get it.	What's an arctan?	"e" for this equation is equal to one because it is a black body ... hope this helps.
Procedural		Use this formula: T(final) = (m1c1T1+m2c2T2) / (m1c1+m2c2). Convert temp to Kelvin and then for your final answer convert back to Cel.	Thanks, I just realized it. I was supposed to solve for cos(c) by moving everything to the other side of the equation then take the cosˆ-1 of that.	Use equation for torque: torque = current * area * sin(90) It is 90 because it is a rectangle. Once you solve for torque multiply it by the N they give you and that is your answer. Make sure to convert your mA to A and cm to m before putting into equation.
Conceptual		I thought you could use the equations for rolling without slipping ... can anyone clarify as to why not?	Do not add 90 degrees. Your answer depends on which quadrant your angle is in. You want the answer to be in the upper right quadrant, so add 180 to the absolute value of your answer if you have a negative x component value to find the angle you are looking for.	I have the correct answer, but I don't understand why it is correct. Why would there be an acceleration at the ball's highest point? Why wouldn't it be zero?

and targeted, the majority fell clearly into one of the classes. If a longer contribution had aspects of more than one class, it was characterized by the class that its majority fell into. However, in a future study, the design should likely allow for more than one classification, such that each contribution can have fractional membership in more than one class. Reliability and generalizability of the classification could be enhanced by asking more than one instructor to classify each contribution, and being able to fractionally consider each judgement in case of disagreements.

Discussion contributions by teaching assistants and instructors were not considered. Also, the correctness of the posting was not considered, e.g., a discussion entry was considered "conceptual" even if it drew the wrong conclusions. Table III shows the distribution of the available discussion contributions.

TABLE III: Classification of the online discussion contributions according the classification scheme described in subsection II B. The columns denote the different discussion types and subtypes, while the rows denote the features.

	Emotional		Surface		Procedural		Conceptual		
	Pos	Neg	Q	A	Q	A	Q	A	
Unrelated	71	54	10	1			1		137
Solution	279	185	601	341	353	456	12	3	2230
Math	1	6	49	36	73	87	3	6	261
Physics		14	85	81	170	190	100	126	766
	351	259	745	459	596	733	116	135	3394

Different classes were combined into the following "superclasses":

Chat - all contributions that are unrelated or emotional.

Emotional climate - the number of positive non-unrelated contributions minus the number of negative non-unrelated contributions. This number would be negative if the problem led to mostly negative emotional comments.

Type and feature sums - number of all related contributions belonging to a certain type, subtype, or feature.

A discussion contribution can be in more than one superclass, for example both "Chat" and "Physics-Related." Figure 4 shows an example of a homework problem and its associated discussion, as well as the appropriate discussion entry classification. The majority of the discussion contributions were of type surface-level or procedural, followed by emotional contributions. The vast majority of discussion contributions had the feature of being solution-oriented, yet a considerable number dealt with the physics of the problems.

III. RESULTS OF ANALYSIS BY STUDENT CHARACTERISTICS

A. Participation

Within the first semester of the calculus-based course, an analysis by student characteristics was performed. Table IV shows the equivalent of Table III for this subset of the data. Out of the 211 students who completed the course, 138 students (65 percent) contributed at least one discussion posting over the course of the semester. Figure 5 shows the distribution of number of discussion contributions over the course of the semester. Most students who participated made between one and ten contributions, but one student made 66 postings. It is not possible to find out which percentage of students *read* the discussions, since they are automatically attached to the problems and always visible. The average number of postings per student was 5 ± 0.7; female students contributed an average of 5.9 ± 1 postings, while male students contributed an average of 3.7 ± 0.7 postings.

TABLE IV: Same as Table III for the first semester of the calculus-based class only. The table includes a small number of contributions by students who eventually dropped the course, which were included in the analysis by problem type, but not in the analysis by student characteristics.

	Emotional		Surface		Procedural		Conceptual		
	Pos	Neg	Q	A	Q	A	Q	A	
Unrelated	14	13							27
Solution	68	56	181	58	141	154	6	1	665
Math			14	11	18	38	1	1	83
Physics		2	22	23	70	85	43	58	303
	82	71	217	92	229	277	50	60	1078

B. Grade-Dependence of Discussion Contributions

The average grade in the course was 3.21 ± 0.05, with men and women achieving equally high grades (men: 3.29 ± 0.08; women: 3.17 ± 0.05). In terms of absolute numbers, within statistical errors, students with high and low grades in the course participated equally in the discussions. A positive correlation between the participation in this "moderated" discussion forum and the student course grade, as it was found in [3], could not be confirmed in this study.

While the *number* of postings is uncorrelated to course grade, their *classification* (subsection II B) turns out to be correlated: In this analysis, the percentage of prominence of certain classes or superclasses in students' cumulative contributions over the semester was analyzed, i.e., the percentage of the respective student's discussion contributions across all problems that belonged into a certain class or superclass. Note that the outcome is independent of the absolute number of postings a student made, e.g., the discussion behavior of the student who made 66 contributions is weighed equally to that of a student having made only the average 5 contributions.

For each grade, the individual percentage (relative) prominences of these classes for students with that grade were averaged. Figure 6 shows the outcome of this study by discussion superclass. For example, the figure is to be interpreted this way: within the indicated errors, 55 percent of a 3.0 student's discussion contributions were solution-oriented. The lines represent second-order polynomial fits to the data.

The relative prominence of solution-oriented discussion contributions varies most strongly with grade, from 75 percent for a 2.0 student to 45 percent for a 4.0 student. The relative prominence of physics-related and conceptual discussion contributions on the other hand increases with grade. The relative prominence of procedural discussions does not vary significantly with grades and is consistent with 42 percent prominence across grades and

A bug that has a mass $m_b = 4g$ walks from the center to the edge of a disk that is freely turning at 32 rpm. The disk has a mass of $m_d = 11g$. If the radius of the disk is $R = 29cm$, what is the new rate of spinning in rpm?

Student A (anonymous); female; 4.0 *(Emotional/Negative/Physics; Chat)*:

```
What is that bug doing on a disk? Boo to physics.
```

Student B (named); male; 3.5 *(Procedural/Answer/Physics)*:

```
OHH YEAH

ok this should work it worked for me

Moments of inertia that are important....
OK first the Inertia of the particle is mr^2
and of a disk is .5mr^2

OK and angular momentum is conserved
IW=IWo W=2pi/T

then do this

.5(mass of disk)(radius)^2(2*pi/T original)+ (mass of bug)
(radius of bug=0)^2= (.5(mass of disk)(radius)^2(2pi/T))+
(mass of bug)(radius of bug)^2(2*pi/T)

and solve for T
```

Student C (named); female; 3.0 *(Surface/Question/Solution)*:

```
What is T exactly? And do I have to do anything to it to
get the final RPM?
```

Student B (named); male; 3.5 *(Procedural/Answer/Solution)*:

```
ok so T is the period... and apparently it works for some
and not others.... try to cancel out some of the things
that are found on both sides of the equation to get a
better equation that has less numbers in it
```

Student D (anonymous); female; 3.5 *(Procedural/Question/Solution)*:
what did I do wrong?

```
This is what I did. initial inertia x initial angular velocity = final
inertia x final angular velocity. I=mr^2, angular velocity = w... so
my I initial was (10g)(24 cm^2) and w=28 rpm. The number
calculated was 161280 g *cm^2. Then I divided by final inertia to
solve for the final angular speed. I found final Inertia by
( 10g +2g)(24 cm^2)=6912. I then found the new angular speed to
be 23.3 rpm. This was wrong...what did I do incorrectly?
```

Student E (anonymous); male; 4.0 *(Procedural/Answer/Solution)*:
Re: what did I do wrong?

```
im not totally sure since i too got it wrong, but i know
your units are not in SI so that may be part of the problem.
```

Student E (anonymous); male; 4.0 *(Surface/Question/Physics)*:
whats the bug?

```
a particle, a disk, initially part of the big disk, what? a
"bug" doesn't explain what we should consider it, inertia-
wise.
```

Student E (anonymous); male; 4.0 *(Procedural/Answer/Physics)*:
Re: whats the bug?

```
nevermind i got it. initially, the bug has no inertia since
distance from center=0. but at the end when bug is at the
edge of disk, just use I(bug)=mr^2.
```

Student F (anonymous); female; 2.5 *(Procedural/Question/Solution)*:

```
Ok- So I used to formula I initial = mr^2+ (1/2)cm^2 the
bug is in the center so there is no inertia. For the I
final I used to equation = mr^2 + (1/2)cm^2 + ( I of bug)
mr^2
Can someone explain what I am doing wrong?
```

Student A (anonymous); female; 4.0 *(Procedural/Answer/Solution)*: **finally**

```
so i finally got this somehow...
We know that the Iw initial = Iw final.
Iw initial: I = 0.5*mr^2 w= ##rpm*(2pi/60)= # radians/sec
---there is no inertial moment for the bug here, r = 0
Iw final: I' = (I of disk, see above) + m(bug)*r^2
w' = ?
Iw = I'w' <-- solve for w'
My trouble was in converting w' back to rpm, but you just
do:
w'*(60/2pi). There you have it. hope this works, because
i hate this problem more than i could ever say. but i
still love gerd.
```

Student G (anonymous); female; 2.5 *(Surface/Question/Solution)*: **Re: finally**

```
I'm still confused, which r do you put 0 in for? either
spot I either get 0 or the rpm i started wtih and neither
are right... something isn't right, can someone help me
```

Student H (named); male; 3.5 *(Emotional/Negative/Solution; Chat)*: **:sigh:**

```
Wow. So, many, little things, can go wrong in calculating
this. Be careful.
```

Student I (anonymous); female; 3.0 *(Surface/Question/Solution)*: **question?**

```
Everything seems to make sense up to where people say to
put in the radius of the bug.. what would it be? For I
final you do I initial + (mass of the bug) * radius of the
bug) ^2 * w. And you are supposed to solve for w. But what
would the radius of the bug be. Because if it is zero
again i get the same w as before and thats not correct???
```

Student J (anonymous); female; 2.5 *(Surface/Question/Solution)*: **Re: question?**

```
That's the same thing that is messing me up. How is the
answer any different from the initial if the radius of the
bug is zero?
```

Student J (anonymous); female; 2.5 *(Emotional/Negative/Solution; Chat)*: **Re: Re: question?**

```
HELP PLEASE!!!!!
Nothing is working!
```

Student K (anonymous); female; 2.5 *(Procedural/Answer/Solution)*: **Re: Re: Re: question?**

```
for the radius of the bug it is the same as the other
radius of the disk.
```

FIG. 4: Example of a standard numerical homework problem and associated discussion

gender, except for the 23 female 4.0 students, where it is 68±7 percent — the 22 male 4.0 students, by comparison, average 34 ± 8 percent procedural discussions.

Except for the exceptionally high prominence of procedural discussion among the best female students, the results are not surprising, but verify the validity of the classification approach.

At the same time, the results confirm that conceptual

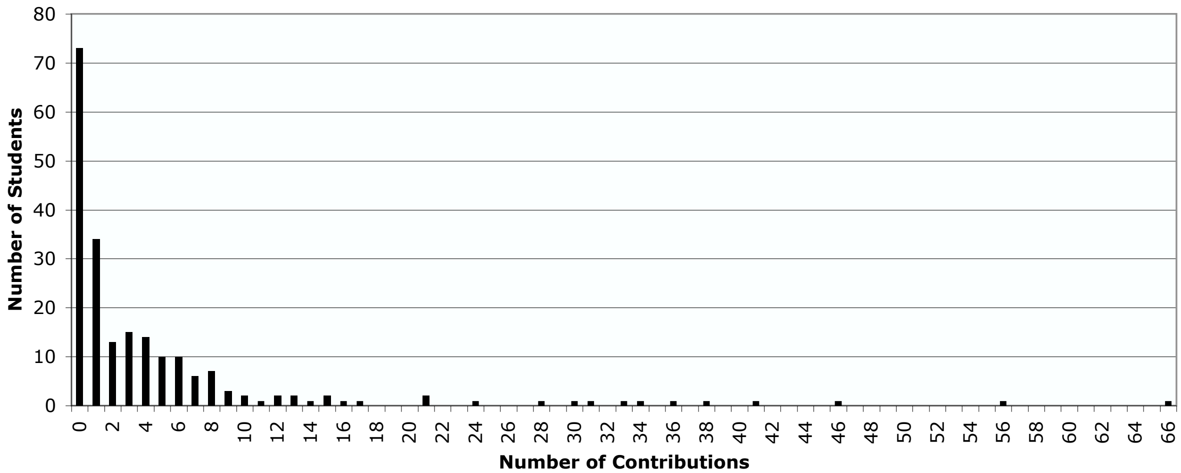


FIG. 5: Number of students versus number of discussion contributions.

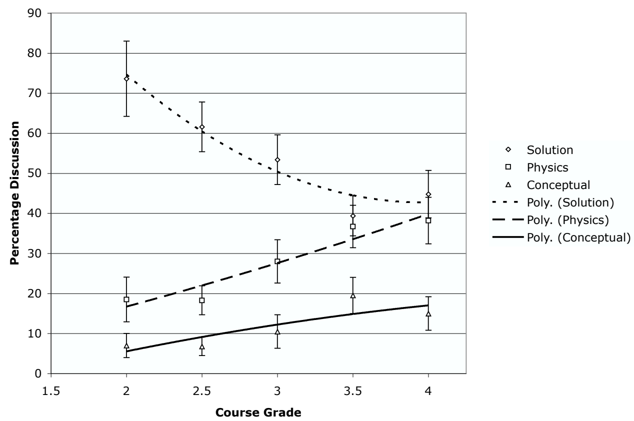


FIG. 6: Prominence of discussion superclasses by grade.

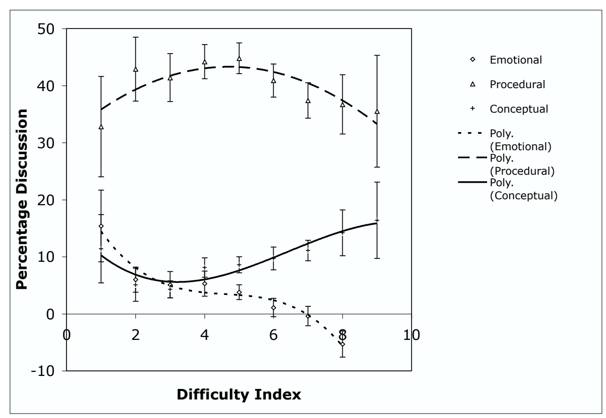


FIG. 7: Discussion characteristics as a function of problem difficulty.

and physics-related discussions are positively correlated with success in the course, while solution-oriented discussion contributions are strongly negatively correlated. While cause and effect may be arguable, in the following section IV, particular attention needs to be paid to problem properties that elicit either the desirable or undesirable discussion behavioral patterns.

Due to the smaller sample size, a correlation analysis by the individual "question" and "answer" classes yielded no statistically significant results.

IV. RESULTS OF ANALYSIS BY PROBLEM CHARACTERISTICS

A. Influence of Problem Difficulty

Using the full data set of all three courses, each discussion contribution associated with a problem was classified according to subsection II B. As a measure of the prominence of a class in a given discussion, the number of contributions belonging to it is divided by the total number of contributions. The discussion characteristics of the problems were binned by their difficulty index and the average percentage plotted in figure 7. Only superclasses are shown (subsection II A), namely the emotional climate (crosses), as well as all (questions and answers) related procedural (triangles) and conceptual (diamonds) contributions. As an example, the plot is to be interpreted in the following way: within the given error boundaries, for a problem with difficulty index of six, ten percent of the online discussion is conceptual. In addition, the data was fit using second order (procedural, long dashes) and third order (emotional climate, short dashes; conceptual, solid) polynomials.

The greatest variation is found in the emotional cli-

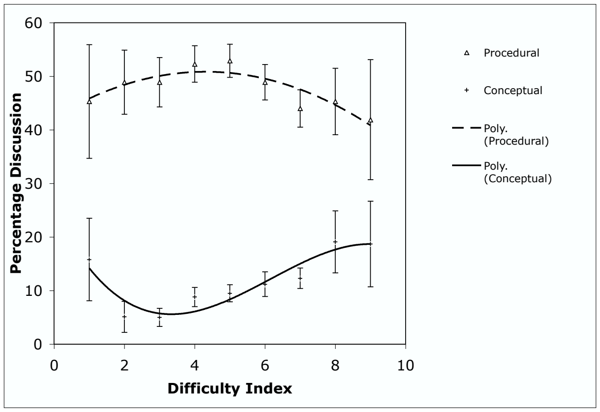


FIG. 8: Discussion characteristics as a function of problem difficulty, no considering "chat."

mate of the discussion. As is to be expected, the climate is mostly positive for "easy" problems, but then remains positive for a fairly wide range of problem difficulties until it becomes negative at a difficulty index of 7. Only six problems had a difficulty index of 9, and — surprisingly — none of these had associated emotional comments.

For difficulty indexes beyond 3, the prominence of conceptual discussions increases. Surprisingly, it also increases for easier problems. This may be attributed to students feeling more confident discussing easier problems on a conceptual level, or simply in there being less of a need of procedural discussions. Overall, the prominence of conceptual discussions is disappointingly low, as it varies between 5 and 16 percent.

Beyond a difficulty index of 5, within error boundaries, the prominence of conceptual discussions would be consistent with a constant 10 percent. If fostering them is a goal, and the emotional climate an indicator of "pain," then beyond a difficulty index of 5 a significant increase in "pain" results in a non-significant gain.

Across all difficulties, procedural contributions dominate the discussions, with relatively little significant variance around the 40 percent mark. The maximum occurs for problems with a difficulty index of 5.

In figure 8 the same analysis was carried out, but this time excluding all "chat" contributions (subsection II A), i.e., only related non-emotional contributions were considered. The relative prominence of procedural and conceptual discussions systematically increases, but all observations from the full analysis remain valid. "Chat" mostly provides a constant background across all difficulty indexes.

B. Influence of Problem Types

Using the full data set of all three courses, each problem was classified according to subsection II A, and each associated discussion entry according to II B. As a measure of the prominence of a class in a given discussion, the number of contributions belonging to it is divided by the total number of contributions. Table V shows the percentage prominence of discussion contributions with a certain type or with certain features in the discussions associated with problems that are of a certain type or have certain features. Error boundaries on the emotional climate values are rather large and mostly include zero (neutral), indicating no significant preferences within the limited sample. Yet, students clearly dislike multiple-choice problems, while they clearly like numerical answer problems. The data also indicates that students prefer "conventional" over representation-translation problems.

The prominence of procedural discussions is significantly higher for numerical problems than for any other problem types, and higher for "conventional" than for representation-translation problems. The latter difference vanishes when "chat" is excluded.

Solution-oriented contributions are significantly higher for multiple-choice and multiple-choice-multiple-response problems than for the other problem types with the exception of formula-response problems, where error-boundaries overlap. In spite of the randomization provided, in discussion entries, students frequently reverse-engineered the complete randomization space by copying their correct answer screens into the discussions (see the example for a surface-level solution-oriented discussion entry in Table II).

The prominence of mathematical discussion contributions is the highest for formula-response problems, approximately equal for numerical and single-response multiple-choice problems, and the lowest for multiple-choice-multiple-response, ranking, and click-on-image problems.

The prominence of physics-related discussion contributions was the highest for ranking and click-on-image problems, and the lowest for multiple-choice problems.

Finally, when it comes to conceptual discussions, their prominence is significantly lower in single-response multiple-choice and numerical problems than in the other problem types. In the earlier study by Kashy [4], it was also found that mastery of these same problem types does not predict overall performance on the final exam as well as other problem types. Multiple-choice problems that do not involve numbers are frequently called "conceptual" problems, but in this study, it was found that they do not necessarily lead to conceptual discussions.

It is a surprising result that the only significant difference between "conventional" and representation-translation problems is that students discuss slightly less procedure in favor of more complaints, and that differences disappear when "chat" is excluded from the analysis. McDermott [7] and Beichner [8] on the other hand

TABLE V: Influence of problem types and features on discussions. The values indicate the percentage prominence of the discussion superclasses, types, and features (columns) for discussions associated with problems of a certain type or with certain features (rows). The values in brackets result from an analysis with "chat" excluded.

	Emot. Clim.	Procedural	Solution	Math	Physics	Conceptual
Multiple Choice	-5±3	28±7 (29±8)	66±7 (74±7)	9±6 (9±6)	16±5 (17±5)	6±3 (7±3)
Short Textual						
Numerical	4±1	48±1 (57±1)	52±1 (63±2)	8±1 (9±1)	23±1 (27±1)	7±1 (8±1)
Formula	6±8	29±11 (31±10)	57±16 (64±18)	31±16 (36±18)		
Mult.-choice Mult.-resp.	1±1	15±3 (16±3)	66±4 (72±4)	1±1 (2±2)	22±3 (26±3)	14±2 (18±3)
Ranking	2±3	24±11 (26±12)	41±18 (46±20)		52±20 (54±20)	38±18 (39±17)
Click-on-Image	0±9	14±6 (18±8)	53±8 (69±11)	3±3 (5±5)	25±11 (26±11)	22±8 (25±9)
"Conventional"	4±1	42±1 (50±2)	55±1 (65±2)	7±1 (8±1)	23±1 (27±1)	9±1 (10±1)
Rep-Trans	-2±2	37±4 (45±4)	52±3 (63±4)	7±2 (9±2)	23±3 (28±3)	8±2 (10±2)

found that students have unexpected difficulties in translating for example data presented as graphs, so a stronger effect of this feature was expected. In addition, Kashy [4] found that mastery of representation-translation problems is the best predictor of final exam scores, even when controlling for ACT, cumulative GPA, and force-concept inventory pretests. Discussion behavior and final exam performance are clearly different measurements for the influence of problem types and do not necessarily need to correlate, but a connection between individual discussion behavior and performance in the course clearly exists (see subsection III B). It should be noted that the earlier study dealt with a relatively small set of representation-translation problems, some of which involved non-static time-evolving simulations as data-source, while in this study, none of the simulation-based problems were assigned. A future study may need to consider the interpretation of time-evolving simulations as a separate feature, once more problems of this type exist in the resource pool.

C. Influence of the course

Few significant differences could be found between the algebra-based and the calculus-based course:

- discussions in the algebra-based course had a significantly higher emotional climate (6±1 versus 2±1)
- the algebra-based course had a higher prominence of "chat" (21±2% versus 11±1% (first semester) and 14±2% (second semester))
- physics-related discussions were significantly higher in the calculus-based course (28±2% (first semester) and 23±2% (second semester)) versus 17±2% in the algebra-based course.
- conceptual-discussions were significantly higher in the first semester of the calculus-based course (12±2% (calculus, first semester) versus 6±2% (algebra)), but this difference vanished in the second semester (7±1% (calculus, second semester)).

Especially the last observation is discouraging, since as the students in the calculus-based course progressed further into their study of physics, the degree to which they were discussing concepts decreased. This might partly be due to the different subject matter (electricity and magnetism versus mechanics), but also due to the lack of reward for conceptual considerations in solving standard homework problems [10].

Again, due to the smaller sample size, a correlation analysis by the individual "question" and "answer" classes yielded no statistically significant results.

D. Qualitative Observations

Reading the online discussions associated with the homework provides valuable insights to the instructor, which are hard to quantify. When assigning homework, instructors usually have an instructional goal in mind, for example, they would like the students to grapple with a certain concept or work through a specific strategy of problem solving. Until the "reality check," the fact that a specific problem only serves this purpose when being approached with an expert mindset is under-appreciated.

An even deeper misconception is the assumption that solving the problem correctly is a reliable indicator of the concept or problem solving strategy being successfully communicated. What the (expert) instructor had in mind, and what the (novice) learner actually does, can be worlds apart [10, 11]. Students are going through reasoning processes and steps that are hardly imaginable to the instructor, and more often than not do several times more work than necessary. The situation that they get a problem right for the wrong reasons is rare, but the instances that they get the problem correct with the same (minimal) amount of steps that an expert would are equally rare — in the end, the concept that was meant to be communicated is lost, since due to their approach, the students "don't see the forest for the trees."

As an example, consider the example Figure 4: there is no external torque, and the problem was meant as a simple example of angular momentum conservation.

Since the disk has several centimeters radius, a bug can safely be approximated as a point mass. It is $(\frac{1}{2}m_d r^2 + m_b 0^2)\omega_0 = (\frac{1}{2}m_d r^2 + m_b r^2)\omega$, and therefore $\omega = \omega_0 m_d/(m_d + 2m_b)$. As long as the disk is much larger than the bug, the result is independent of its radius, and no unit conversions are needed. Several things jump out to the expert reader of the discussion:

- No student mentions the fact that there is no external torque or explicitly mentions angular momentum conservation as the starting point for their considerations.
- The idea that a bug could be approximated as a "point mass" compared to the size of the disk is never mentioned, even though Student E raises the issue.
- Regarding the calculation of the moment of inertia, there is confusion between the radius of an extended symmetrical object and the radius of the orbit of a point mass (thus, presumably, the question "what is the radius of the bug?").
- Students are plugging in numbers early and do not eliminate the radius of the disk from their calculations (with the possible exception of Student B who hints that "cancel out some of the things that are found on both sides of the equation to get a better equation that has less numbers in it.").
- Students do not appear to realize that unit conversions are in fact not needed.
- No student simply posts the final symbolic solution, which is true for virtually all analyzed discussions.
- Students went through considerable effort to solve this rather straightforward problem and do not realize that the solution is much simpler to achieve. Note in particular Student H's comment that "so many little things can go wrong." Here, numerical online homework clearly falls short of handgraded homework, since the students are only graded on the correct final solution, not on their solution strategy.

Particularly the last point is distressing, since it instills a false sense of mastery among the students and confirms them in their undesirable techniques, which is an observation already pointed out by Pascarella [12] in an earlier study of online homework systems. The discussion in Figure 4 is typical, in spite of the fact that in lecture, problem solving strategies had been discussed, and examples had been given how the derivation of a final result in symbolic form can lead to faster and more reliable results. When discussing examples during lectures, the instructor attempted to model good problem solving strategies.

Many of these shortcomings may be correctable through early detection, and closely following the online student discussions prior to lecture, particularly around the assigned reading problems, may be a valid extension of the Just-in-Time Teaching [13] technique.

E. Comparison to other research approaches

The presented method to gain insight into student problem solving behavior is comparable to the more traditional "thinking out loud" or group discussion observations. However, in the former method, the subjects are keenly aware of the observer, which may influence their behavior: in most any course, appropriate problem solving techniques would have been discussed, and while in reality, students might find them "inefficient" or "slow" [10], they might try hard to exhibit them in the research setting. The latter method, observation of student discussions, is likely closer to the behavior students would exhibit when not observed. However, groups are smaller, and in most studies interact around problems less complex than the average homework problem.

An advantage for the researcher of this method is the ready availability of the online discussions — there is no need for transcription, since the discussions are already in textual form. In addition, since written student discussions contain less spurious verbiage and slang, and tend to exhibit better grammar and more complete sentences than the spoken word, evaluation is easier. The discussion contributions are likely to closely reflect students' actual approach, since students would aim to solve the homework in the way they believe is most efficient. The large number of discussion contributions allows for statistically significant results. A disadvantage is that this method depends on problem randomization, and thus can only be used with systems like LON-CAPA. Were the online problems not randomizing, discussions would likely consist of one or two entries only with the final answer, such as "17.5 m/s" or "Answer B." Also, the online system must not have a separate discussion area, but provide contextual discussion functionality.

In the current study, a general classification scheme was deployed across physics topics and concepts. However, since the discussions are associated with certain problems, they can also be used to study student understanding of certain topics.

V. CONCLUSIONS

Online student discussions are a rich source of insight into student problem solving behavior. It was verified that indeed conceptual and physics-related discussion contributions are characteristics of students who are successful in the course, while the prominence of solution-oriented discussion contributions is strongly negatively correlated with success in the course.

Different discussion patterns ensue around different problem characteristics:

Difficulty Very easy problems can elicit a high level conceptual discussions, and so can problems of midrange difficulty. As problems become more difficult, there is no significant gain in conceptual discussions.

Problem Types Different problem types result in different associated discussion patterns. Discussions on a procedural level are more prominent for numerical problems than for any other problem type. Solution-oriented discussions are more prominent for multiple-choice style problems in an effort to short-circuit the conceptual reasoning. Discussions around single-response multiple choice problems and numerical problems have a significantly lower prominence of conceptual discussions than other problem types. Ranking problems show very favorable discussion patterns, but their sample size has been too small to make definitive statements.

Analyzing online discussions around problems has been found to provide valuable insights into student problem-solving strategies.

VI. OUTLOOK

In this current study, little is known about the students except their gender and final course grade, and the analysis of discussion behavior by student characteristics thus yielded less results than the study by problem characteristics. Research [14] suggests that learning processes are strongly influenced by epistemological beliefs, and it will be interesting to analyze the correlation between attitudes and beliefs (as measured for example by the MPEX [15]) regarding physics and online discussion behavior. Also, the final grade in the course incorporates a lot of factors including some measures of simply diligence, and interactions with for example the FCI [16] gains might result in better correlations to the students' conceptual understanding of physics.

Reliable and consistent coding of discussion contributions is crucial for the analysis. Any future study should include more than one instructor in coding the discussion contributions to increase reliability and generalizability of the results, as well as to avoid possible personal bias. As the online material in the courses changes, more and more numerical problems are replaced or enhanced by other problems types, which hopefully will yield more statistically significant results.

Acknowledgments

Supported by the National Science Foundation under NSF-ITR 0085921 and NSF-CCLI-ASA 0243126. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

---

[1] E. Mazur, *Peer Instruction* (Prentice Hall, 1997), ISBN 0-13-565441-6.
[2] R. Wallace, Education, Communication and Information **3**, 241 (2003).
[3] D. A. Kashy, G. Albertelli, W. Bauer, E. Kashy, and M. Thoennessen, Journal of Asynchronous Learning Networks **7**, 31 (2003).
[4] D. A. Kashy, G. Albertelli, G. Ashkenazi, E. Kashy, H.-K. Ng, and M. Thoennessen, in *Proc. Frontiers in Education* (2001), vol. 31, pp. 1073–1078, ISBN 0-7803-6669-7.
[5] R. N. Steinberg and M. S. Sabella, Phys. Teach. **35**, 150 (1997).
[6] E. F. Redish, *Teaching Physics* (Wiley, 2003), ISBN 0-471-39378-9.
[7] L. McDermott, M. L. Rosenquist, and E. H. van Zee, Am. J. Phys. **55**, 503 (1987).
[8] R. J. Beichner, Am. J. Phys. **62**, 750 (1994).
[9] E. Mazur, Optics and Photonics News **6**, 59 (1996).
[10] H. Lin, Phys. Teach. **20**, 151 (1982).
[11] M. T. H. Chi, P. J. Feltovich, and R. Glaser, Cognitive Science **5**, 121 (1981).
[12] A. Pascarella, in *Proc. NARST Annual Meeting* (2004).
[13] G. M. Novak, E. T. Patterson, A. D. Gavrin, and W. Christian, *Just-in-time teaching: blending active learning with web technology* (Prentice Hall, 1999), ISBN 0-13-085034-9.
[14] V. Aleven, E. Stahl, S. Schworm, F. Fischer, and R. Wallace, Review of Education Research **73**, 277 (2003).
[15] E. F. Redish, R. N. Steinberg, and J. M. Saul, Am. J. Phys. **66**, 212 (1998).
[16] I. Halloun, R. R. Hake, E. P. Mosca, and D. Hestenes, *Force concept inventory*, URL `http://modeling.la.asu.edu/R&E/Research.html`.