

Correlations among knowledge structures, force concept inventory, and problem-solving behaviors

Kathy L. Malone

Shady Side Academy, 423 Fox Chapel Road, Pittsburgh, Pennsylvania 15238, USA

(Received 6 July 2007; published 20 November 2008)

The modeling instruction pedagogy for the teaching of physics has been proven to be quite effective at increasing the conceptual understanding and problem-solving abilities of students to a much greater extent than that of nonmodeling students. Little research has been conducted concerning the cognitive and metacognitive skills that modeling students develop that allow for these increases. Two studies were designed to answer the following question: In what ways do the knowledge structures, metacognitive skills, and problem-solving abilities differ between modeling and nonmodeling students? In study 1, the knowledge structures developed by two groups of high school physics students taught using differing pedagogies (modeling instruction in physics and traditional methods) were determined using a card-sort task. The student's knowledge structures were then correlated with the scores they obtained on two measures: the force concept inventory (FCI) and a problem-solving task (PS task) developed for this study. The modeling students had a more expertlike knowledge structure, while the nonmodeling students produced structures that were novicelike. In addition, the expert score correlated highly with performance on both the FCI and PS task scores demonstrating that a higher expert score predicted a higher value on each of these measures while a higher surface feature score predicted a lower score on both of these measures. In study 2, a verbal protocol design allowed for a detailed study of the problem-solving and metacognitive skills utilized by the two groups. It was determined that the skills utilized by the modeling instruction students were more expertlike. In addition, the modeling students produced significantly fewer physics errors while catching and repairing a greater percentage of their errors.

DOI: [10.1103/PhysRevSTPER.4.020107](https://doi.org/10.1103/PhysRevSTPER.4.020107)

PACS number(s): 01.40.Fk, 01.40.Di

I. INTRODUCTION

Modeling instruction in physics is a high school physics pedagogy that is built on the premise that active engagement is necessary and as such it is constructivist based.¹ The efficacy of modeling instruction has been demonstrated in a number of studies using the force concept inventory (FCI) and the mechanics baseline test (MBT).²⁻⁵ Prior research has shown that a stronger knowledge base as measured by the FCI is correlated with problem-solving performance as measured by the MBT.²⁻⁴ These positive results led the Department of Education to recognize the pedagogy as an exemplary program in science education in 2001.

In modeling instruction students are taught to organize knowledge in terms of basic physics models (e.g., constant velocity, constant acceleration, etc.) developed by the students using data obtained in paradigm laboratories. Different representations (graphical, verbal, algebraic, and diagrammatic) are associated with the models during development. For example, the algebraic representation for the constant velocity model would be

$$\vec{x}_f = \vec{v}\Delta t + \vec{x}_i.$$

The graphical representation would include position time and velocity time graphs, while the diagrammatic representation would include motion maps. As shown in Fig. 1, a motion map shows the location and motion of an object at specific time periods. In addition, students are taught to associate appropriate solution procedures to each model. This coherent organization should allow students to easily identify classes of physics problems and to apply the associated solution procedures. It was demonstrated that novice students' knowl-

edge was fragmented and poorly organized around surface features, while an expert's knowledge was organized around principles.^{6,7} It is possible that the coherent organization developed by the modeling students allows them to demonstrate more expertlike problem-solving skills. Evidence for this assertion that the structure of one's knowledge may play an important role was uncovered when Hinsley *et al.*⁸ demonstrated that competent problem solvers in algebra did indeed utilize schemas and that these schemas seemed to direct their problem-solving strategy. A schema is a mental structure that allows one to organize their knowledge. For example, in math a problem schema would consist of interrelated solution methods for particular types of problems that unite the problems on the basis of some type of underlying feature or features. Hinsley *et al.*⁸ asked math students to categorize a set of algebra word problems by problem type. They found that students did categorize problems into type and this categorization occurred very quickly sometimes after reading only the first statement in the problem. They went on to explore if the students used these categories to solve problems. They discovered that they did indeed utilize them to help solve problems and that the categories included information about "useful equations and diagrams and appropriate procedures for making relevant judgments."⁸ In addition, the physics education research group at the University of Massachusetts in a series of experiments showed that novice students taught a more expertlike knowledge structure



FIG. 1. Motion map of an object moving at a constant positive velocity.

demonstrated a 15% increase in problem-solving ability over control students.^{9–11}

Modeling instruction requires students to account for their decisions by explaining how they know what they know, and whether their answer or approach makes sense. In this way, modeling instruction encourages students to develop metacognitive problem-solving skills to consider what they do and to evaluate their final conclusions. Schoenfeld¹² designed a college problem-solving course which contained a large metacognitive component such that the decision making processes were role modeled during problem solving. As the students solved problems they were always asked “Why are you doing that?” and “How does it help you?.” Questions such as these ask the students to monitor their comprehension and explicitly evaluate their progress toward the solution. Schoenfeld¹² analyzed video taped student solutions precourse and postcourse and discovered that the students were using a greater number of metacognitive skills thereby demonstrating more control of the solution path and performed more expertly postcourse. Schoenfeld¹² also demonstrated that the students in the course performed significantly better than a control group on problem-solving tasks. A similar strategy to teach metacognitive skills to middle school students called improve has produced improved problem-solving performance on quantitative post-tests.¹³ Research results such as these in cognitive science would suggest that the use of such skills should allow the modeling students to catch and correct more physics errors.

While no studies to date have attempted to specifically analyze the cognitive structures developed by modeling students and how those structures might help during problem solving, Brewster¹⁴ determined indirectly that his modeling-oriented energy-thread curriculum seemed to enhance the development of expertlike students. When solving complex paper and pencil problems Brewster’s scoring rubric¹⁴ determined that the modeling students represented and analyzed physics problems to a greater degree than nonmodeling students. A strong characteristic of expert physicists’ problem-solving strategies is the increased use of problem representations. Students with higher scores for representation on complex problems were shown to be more likely to correctly answer the problems. Thus, Brewster¹⁴ believed that his curriculum aided students in creating a better organized knowledge base. The purpose of the studies presented here is to more directly attempt to analyze the cognitive structures developed by modeling students and to determine how these structures aid in problem solving.

II. STUDY 1: HOW DOES MODELING INSTRUCTION INFLUENCE STUDENTS’ COGNITIVE STRUCTURES?

A. Experimental design

This study consisted of two classes of first year high school students ($n=61$) taught physics via the modeling instruction pedagogy by two different instructors (one of which was the author) and two classes of first year students ($n=36$) taught via traditional methods by two different instructors. The teachers involved in the study were veteran instruc-

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1									
1.2									
1.3									
2.1									
2.2									
2.3									
3.1									
3.2									
3.3									

FIG. 2. Example of expert matrix used for card sort.

tors of physics. The two modeling teachers had been modeling for 13 years. One of the modeling teachers had been teaching for 30 years, while the other had been teaching for 20 years. One of the nonmodeling teachers had been teaching for 30 years, while the other had been teaching for 13 years. All of the schools participating in this study were suburban with similar socioeconomic backgrounds (majority middle to upper class) that support close to 100% of all graduates moving on to postsecondary training, primarily four-year degree programs. All of the schools in the study have similar length school days and school years.

The knowledge organization of the students was assessed using a card-sort task similar to that utilized in past research.^{6,8} The card-sort task consisted of six sets of problems written on cards. Each set required the use of a single model (i.e., constant velocity, constant acceleration, Newton’s second law, impulse and momentum, circular motion, or conservation of energy) but each problem was based on a different cover story or surface feature (i.e., incline plane, springs, etc.). The students were asked not to solve but to place the problems into groups based on the similarity in solution strategies. Each of these models would be developed in a nonmodeling class but would not be called models and would be developed in a piecemeal fashion. Appendix A contains an example of one isomorphic problem set.

To compare the knowledge structures developed by each group of students, a confusion matrix was used to obtain expert and surface feature scores (hereafter referred to as the SF score) from the card-sort groupings. One matrix was based on an expert model grouping and a second was based on surface feature groupings. Therefore, each problem was assigned a number with the ones digit based on the deep structure model used and tenths digit based on the surface feature addressed by the problem. For example, problem 2.2 was given this value because its deep structure consisted of constant acceleration while its surface feature was based on an inclined plane. Whereas, 3.2 would be considered a constant force model problem (i.e., Newton’s second law) using a cover story of an inclined plane. An expert modeler would sort all of the cards solved with the deployment of the constant velocity model together, all of the constant acceleration model problems together, etc. Therefore, an expert confusion matrix, as seen in Fig. 2, would have all of the cards solved using the constant velocity model grouped together and entered into the square with all of the problems starting with a

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1		X			X				X
1.2	X				X				X
1.3									
2.1									
2.2	X	X							X
2.3									
3.1									
3.2									
3.3	X	X			X				

FIG. 3. Example matrix with markings for a sort that grouped problems 1.1, 1.2, 2.2, and 3.3.

ones digit. An X was placed in the column under problem 1.1 next to each problem that a student assigned to that category. For example, if a student had sorted problems 1.1, 1.2, 2.2, and 3.3 together there would have been an X placed in the boxes in the column under problem 1.1 at the location of 1.2, 2.2, and 3.3 as shown in Fig. 3. An expert or deep structure sorting would have the majority of the groupings falling into the bordered boxes. The expert score was obtained by dividing the number of hits in the expert boxes vs the number of hits in and out of the expert boxes (then multiplying by 100). Therefore, a student with a card sort that was based entirely on models would obtain a value of 100, whereas, a student sorting based strictly on surface features would receive a score of zero. A sample expert score calculation is shown in Fig. 4.

A companion SF score was determined in a similar fashion. A surface feature matrix was completed for each student. The main difference between this matrix and the expert matrix is that all the problems with the same tenths digit were grouped together instead of the ones digit. The SF percentage score was calculated in a similar fashion as the expert score as shown in Fig. 5.

During analysis it was discovered that students were also sorting problems based on what the question was asking them to solve for (referred to as the questions-asked strategy or QA strategy). This was unexpected as it was not mentioned in past research. However, two pieces of research conducted in junior high math and college physics classes demonstrated that students did sort problems via the question asked.^{15,16} Upon further review, the study of Chi *et al.*⁶ did mention that novices do use literal features of the problem during card sorts. The questions-asked strategy could be con-

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1		X	X				X		
1.2	X		X				X		
1.3	X	X					X		
2.1					X	X			
2.2				X		X			
2.3				X	X				
3.1	X	X	X						
3.2									X
3.3								X	

Expert score = ((number of hits in the expert boxes)/(total number of hits in and out of expert boxes)) * 100
 Expert score = (14 / (14+6)) * 100 = 70 %

FIG. 4. Example of an expertlike card sort and the calculation of the score.

	1.1	2.1	3.1	1.2	2.2	3.2	1.3	2.3	3.3
1.1		X	X						
2.1	X		X						
3.1	X	X							
1.2					X	X		X	
2.2				X		X		X	
3.2				X	X			X	
1.3									X
2.3				X	X	X			
3.3							X		

Surface feature score = ((number of hits in the boxes)/(total number of hits in and out of boxes)) * 100
 Surface feature score = (14 / (14+6)) * 100 = 70 %

FIG. 5. An example of a surface feature matrix for a novice student sorting cards based mostly upon surface features.

sidered a literal feature. Therefore, a third matrix was developed based on the questions asked in each problem. A sample questions-asked matrix is shown in Fig. 6. The questions-asked score (hereafter referred to as the QA score) was determined for the questions-asked matrix using the same formula developed for expert and SF scores. There were a total of seven different “types” of questions asked among all of the problems. The percent of expert groupings that can be explained by the questions-asked groupings ranges from a high of 34% to a low of 5% depending on the question asked. For example, the greatest preponderance of questions asked the students to solve for the final velocity, initial velocity, or average velocity. The questions-asked velocity grouping accounted for 34% of the expert model groupings. All of the other questions-asked groups accounted for approximately 17% or less of the expert model groupings which is very close the percentage accounted for by a random grouping (i.e., 14%). Therefore, the data discovered from this matrix must be looked at carefully since the questions are biased in certain directions since the problems were only designed to be isomorphic around models and surface features. However, the results could lead to interesting comparisons between groups.

Conceptual understanding was assessed using the FCI while problem solving was assessed by the problem-solving task (PS Task) designed specifically for this study. The PS task consisted of seven quantitative problem-solving tasks. There was one question each for the six models using in this study: constant velocity, constant acceleration, circular motion, energy, Newton’s second and third laws, and impulsive force. The seventh task was a graphical constant velocity–constant acceleration task. Three of the tasks (circular mo-

	1.1	1.2	1.9	2.5	4.9	1.7	2.1	5.3	6.8
1.1		X	X	X					
1.2	X		X	X					
1.9	X	X		X					
2.5	X	X	X						
4.9						X	X		
1.7					X		X		
2.1					X	X			
5.3									X
6.8								X	

FIG. 6. Sample questions-asked matrix example demonstrating the nonuniformity of the questions if grouped purely on the questions asked.

TABLE I. Study 1: test scores for modeling and nonmodeling groups

Post-test scores	Nonmodeling classes		Modeling classes			Statistics between groups	
	First-year post-test— class 1	First-year post-test— class 2	First-year pretest— class 1	First-year post-test— class 1	First-year post-test— class 2	<i>F</i> statistic	<i>p</i> value
FCI	39	68	27	65	65	31.81 (C1) 0.32 (C2)	$p < 0.001$ NS
Expert card sort	30	33	28	44	41	18.74	$p < 0.001$
SF card sort	51	30	19	18	21	82.17 (C1) 9.2 (C2)	$p < 0.001$ $p < 0.003$
Question asked card sort	23	33	46	34	40	3.11	$p < 0.081$
PS task	42	50		62		10.76	$p < 0.002$

tion, Newton's second law, and energy) were adapted from the mechanics baseline test with permission from Hestenes and Wells.³ The problems were designed so that all of the major models covered in the modeling curriculum were included and based on problems that all physics textbooks traditionally cover. The graphical task was added so that it could be determined if participants in both conditions could solve problems utilizing graphical methods in addition to algebraic methods. The PS task was given within two weeks of completing all of the traditional kinematics, mechanics, and energy concepts in the course currently taken by the students. The internal consistency of the PS task was assessed using the Kuder-Richardson formula 20 (KR-20). The calculated KR-20 reliability score of 0.76 is an adequate value for use in making comparisons.¹⁷ However, obtaining a score lower than 0.80 was not surprising given the small number of items on the test due to the time constraints of a 40 min high school class.¹⁸ If the Spearman-Brown prophecy formula for 14 items is applied the estimated KR-20 value would increase to 0.87.¹⁹ The PS task in its entirety is located in Appendix B.

In addition, modeling students were given two pretests: the card-sort task and the FCI. It was initially planned that the nonmodeling students would also take the same pretests but due to administrative difficulties at these schools we were unable to complete all of the assessment tasks as initially planned.

B. Hypotheses

The following hypotheses were advanced:

(a) Students taught using modeling instruction in physics will display a larger gain on the FCI and a shift to more expertness.

(b) Knowledge structures will be expertlike (based on models) for modeling students and novicelike (based on surface features) for nonmodeling students.

(c) Modelers will outperform nonmodelers on all post-tests.

(d) Test scores will be highly correlated with knowledge structure.

C. Results

Scores and statistics for all comparisons made in study 1 are summarized in Table I. Modeling students demonstrated a gain from precard to postcard-sort scores. There was a shift toward a model-based knowledge structure (a 44.4% relative difference).²⁰ In addition, the modeling students demonstrated a significant gain in conceptual understanding as shown by the FCI [an 82.6% relative difference²⁰ or a normalized gain of 0.52 (Ref. 21)].

Initially, a comparison of the post-test scores for each group of classes was carried out. An ANOVA revealed significant differences between the two nonmodeling classes: post-test FCI and SF scores. The average FCI score of 68% for the second nonmodeling class was surprising since it did not fall in line with previous research which found traditional nonmodeling classes scored 50% or lower on the FCI.⁴ In addition, past research showed that a high FCI is necessary but not sufficient for a good score on a problem-solving task such as the MBT.³ Therefore, normally one would expect that a higher FCI score would allow for a higher PS task score. However, these data demonstrate that the PS task scores between the two nonmodeling classes are not statistically significant even through the post-FCI scores were. The data imply that the methods used in the second nonmodeling class seemed to elevate the FCI scores while not increasing the participants' problem-solving abilities. An ANOVA revealed no significant differences between the two modeling classes' post-test scores. For further data analysis, all the post-test scores that were not statistically significant within the two groups were collapsed.

A comparison of scores between pedagogies was completed. An ANOVA demonstrated that overall the modelers' knowledge organization, conceptual understanding and problem-solving abilities differed significantly from that of the nonmodelers after a year of physics. The *F* statistics and *p* values for all comparisons made are summarized in Table

TABLE II. Summary of ANOVA results between modeling and nonmodeling class post-test scores. Scores that are statistically significant are shaded. NS stands for nonsignificant.

Comparison	<i>F</i> statistic	<i>p</i> value
Expert card-sort score for collapsed nonmodeling class vs collapsed modeling classes	$F(1,95)=18.74$	$p<0.001$
SF card-sort score for nonmodeling class 1 vs collapsed modeling classes	$F(1,83)=82.17$	$p<0.001$
SF card-sort score for nonmodeling class 2 vs collapsed modeling classes	$F(1,71)=9.2$	$p<0.003$
Questions-asked card-sort score for collapsed nonmodeling class vs collapsed modeling classes	$F(1,95)=3.11$	$p<0.081$
FCI score for nonmodeling class 1 vs collapsed modeling classes	$F(1,83)=31.81$	$p<0.001$
FCI score for nonmodeling class 2 vs collapsed modeling classes	$F(1,71)=0.32$	NS
PS task score for collapsed nonmodeling class vs collapsed modeling classes	$F(1,67)=10.76$	$p<0.002$

II. These results demonstrate that the expert card-sort scores between pedagogies are significant. The modeling group has a significantly higher expert knowledge structure score than the nonmodelers. A Monte Carlo simulation was completed using 10% of the participants selected randomly. It was determined that a random expert post-test score was 15 on average with scores ranging from 12 to 18. None of the post-test scores obtained in this study are near to the random value; therefore, the sort is reflecting the student's conceptions at that time. A qualitative analysis of the card sort showed that a number of students (especially modeling students) were collapsing a number of the models by the end of the school year. For example, a number of students collapsed constant velocity model with constant acceleration model which is not surprising since constant velocity is a special case of zero constant acceleration. This type of collapse seems to be demonstrating a more expertlike view since all of the models could be collapsed into two fundamental principles, i.e., energy and momentum.⁶ However, these types of collapses actually tend to depress the expert score as it is presently being calculated.

Even though the two nonmodeling classes differ on their surface feature dependence by the end of the year it was found that both nonmodeling surface feature post-test scores differed significantly from that of the modeling group. This finding demonstrates that the nonmodeling group as a whole is more surface feature oriented than the modeling group and supports the increased expertlikeness of the modeling group. A Monte Carlo simulation was completed using 10% of the participants selected randomly. It was determined that a random SF post-test score was 13 on average with scores ranging from 10 to 20. In fact, the nonmodeling group's average post-test scores are approaching the random range of scores. The results showed that the nonmodelers' knowledge structures were based on surface features as predicted by past studies^{6,9,22} while the modeler's structure was not (a 70% relative difference²⁰).

The mean card-sort scores recorded in Table I suggest that when the average expert score is larger the SF score on av-

erage is lower. Curve estimation between the expert scores and SF scores was completed using the data obtained from all of the participants. It was determined that the best relationship estimate between the two variables was an inverse ($R=0.63$). The best relationship determined was

$$\text{expert score} = 0.06/\text{SF score}.$$

The questions-asked scores across all groups were marginally significant with the modeling group demonstrating a larger score by the end of the year. A Monte Carlo simulation was completed using 10% of the participants selected randomly. It was determined that a random questions-asked post-test score was 19 on average with scores ranging from 17 to 23. The scores by all groups on average were either outside of this range or in the case of the nonmodeling class 1 equal to the maximum value obtained in the simulation. A larger score on the part of the modeling classes might be caused by the decline in SF score and the attempt on the part of the students to use a more sophisticated strategy while on the way to becoming more "expertlike." It is possible that the use of surface features to sort problems is a novice sorting and that as sophistication increases students might realize that problem solving based on surface features leads to less success thus causing them to develop a questions-asked strategy. The questions-asked sorting could later be modified by students as their knowledge structures become more robustly expertlike with additional learning.

It might be expected that the PS task score for the nonmodeling class 2 to be more comparable to the modeling groups' score given the high FCI score but this is not the case. The PS task scores for all the nonmodeling groups were significantly lower than the modeling groups. These findings clearly demonstrate that the modeling pedagogy allows for the production of a more expertlike knowledge structure to develop which allows the participants to perform better on both qualitative and quantitative tests than traditional students.

A correlation between FCI post-test scores and PS task scores was conducted utilizing the data from all groups. The

TABLE III. Summary of correlation coefficient and R^2 values between all post-tests. Items that are statistically significant are italicized.

Test	Correlation coefficient (R)	R^2 (%)
<i>FCI vs PS task</i>	<i>0.62</i>	<i>38</i>
<i>Expert score vs FCI</i>	<i>0.57</i>	<i>33</i>
<i>Expert score vs PS task</i>	<i>0.65</i>	<i>41</i>
<i>SF score vs FCI</i>	<i>-0.50</i>	<i>25</i>
<i>SF score vs PS task</i>	<i>-0.45</i>	<i>20</i>
<i>SF score vs expert score</i>	<i>-0.50</i>	<i>25</i>
QA score vs FCI	-0.24	6
QA score vs PS task	-0.13	2
QA score vs expert score	0.08	0.06
QA score vs SF score	0.24	6

correlation coefficient between the FCI and the PS task was 0.62. The data are in line with the correlation coefficient value of 0.68 obtained by Hestenes and Wells³ showing that the FCI score accounts for a large proportion of the variance in problem-solving tasks. Correlation tests were performed with all of the data to determine if any of the card-sort scores were predictive of the FCI and/or PS task scores. Table III summarizes these analyses. The correlations are quite high for expert score vs FCI score and expert score vs PS task score. SF or questions-asked scores have a negative correlation for both FCI and PS task scores but only surface feature correlations are significant. Thus, the higher the SF scores the lower the FCI and PS task scores. Correlations were also completed between the different card-sort scores. It was found that the SF score and the expert score were negatively correlated with a correlation value of 0.50.

Since there is no underlying theory concerning what the FCI measures, a stepwise regression was completed to determine if the FCI and expert scores together would be better predictors of the PS task than either. It was demonstrated that this was not the case. It was found that together the expert score and FCI score accounted for 44% of the variance in PS task score. This combination allows one to account for an increase in the variance of only 6%. Rather, the two scores seem to be predicting the same variance in the PS task suggesting that they are testing for similar developments in student cognition.

In addition, the predictiveness of the expert score toward problem-solving ability seems to be less influenced by pedagogy than is the FCI score. This finding was discovered when the correlations between the PS and FCI scores varied greatly when the nonmodeling class 1 data were excluded (0.53–0.65) while there was no such variation in correlations between the expert score and the PS score (0.63 vs 0.64, respectively). Therefore, some instructional methods can overly influence the FCI scores without increasing the problem-solving ability of the students while the corresponding expert score obtained continues to be predictive of problem-solving ability.

III. STUDY 2: HOW DOES MODELING INSTRUCTION INFLUENCE PROBLEM-SOLVING AND METACOGNITION?

A. Research questions

This study was designed to compare the problem-solving strategies used by high school physics students taught utilizing modeling instruction and those taught utilizing nonmodeling pedagogy. The research questions were:

- (1) Will the modeling participants demonstrate more expertlike problem-solving strategies than the nonmodeling students?
- (2) Will there be metacognitive differences (planning, monitoring, and evaluating) between the two groups?
- (3) Will the two groups differ in the detection and correction of physics errors?

B. Experimental design

All of the A and B students from study 1 were invited to volunteer for study 2. The volunteers consisted of 19 modeling and 11 nonmodeling students. A verbal protocol design required students to talk aloud while solving five problems. The series of five physics problems was designed so that each problem required the deployment of a particular physics model. Four of the problems were designed to test the following models: constant acceleration, Newton's second law (constant force), conservation of momentum (impulsive force), and energy. There was an additional problem that utilized two models at once (constant acceleration and constant velocity models) and was graphical in nature. This particular problem was designed so that it was more easily solved via a graphical approach in order to test the students' flexibility with utilizing alternate solution representations. The problems contained confounding information such as unnecessary information, information that suggested alternative models, or graphical information. For example, the following problem was the third verbal protocol problem out of the sequence of five:

A comet has a mass of 50 000 kg and is in an elliptical orbit whose long axis is 6×10^{12} m, and its period of revolution around our sun is 100 years. An asteroid has a mass of 10 000 kg and is in a circular orbit whose diameter is 2.2×10^{12} m, and its period of revolution around our sun is 20 years. The comet has an x component of velocity of -5000 m/s. It collides head on with the asteroid, whose x component of velocity is $+11\,000$ m/s. After the collision the comet has an x component of velocity of $+3000$ m/s. What is the final x component of the velocity of the asteroid?

This problem was designed such that it could activate either the circular motion model or gravitational field model for students while requiring the deployment of the impulsive force model. All of the problems used in the verbal protocol are contained in Appendix C.

Each participant was video taped. The video camera was situated so that it collected not only voice recordings but a

TABLE IV. Comparison of expert and novice problem-solving strategies.

Expert behaviors	Novice behaviors
Typically use a working forward strategy except on more difficult problems	Typically use a working backward strategy
Performs an initial qualitative analysis of the problem situation	Usually manipulates equations discovered via equation hunting
Constructs diagrams during solution process	Rarely constructs or uses diagrams
Spends time planning approach sometimes via models of the physical situation	Rarely plans approach simply dives in
Uses fewer equations to solve the problem	Uses more equations to solve problem
Usually solve problems in less time	Usually takes more time to solve the problems
Refers to the physical principles underlying the problem	Refers to the numeric elements of the problem
Concepts more coherent and linked together	Concepts not coherent and lack applicability conditions for special cases
Fewer errors—concepts usually deployed correctly	More errors—concepts usually deployed incorrectly
Can use more than one representation to solve problems—which usually allows them to deviate to other solution paths when stuck	Usually only utilize a numeric representation to solve problems—once they become stuck rarely can free themselves
Check and evaluate solution by a variety of methods (i.e., more flexible)	Superficially check solution if at all
Rarely refer to problem statement or text	Frequently refer to problem statement and textbook (especially examples)

record of the student's solution path. The students recorded their solutions on large sections of whiteboard using dry erase markers. The participants were allowed to refer to textbooks and notebooks and were given unlimited time but if they seemed to reach an impasse, it was suggested that they move on. After the participants completed the series of five problems a retrospective interview was conducted. The participants were shown the whiteboard they used for each problem and were asked to discuss all that they could remember thinking when solving that problem. Afterward, clarification questions were asked of each subject before moving onto the next problem. A verbal protocol instruction script was used to ensure that all participants were given the same directions. The verbal protocol data were then transcribed and the transcriptions of the audio tapes for each participant were correlated with the video in order to produce a record of the physics path taken by each subject in each problem.

The audio transcription was segmented into chunks corresponding to transitions in the physics problem-solving path. The following transcription segment was taken from a non-modeling subject (information in brackets contains notes made by the transcriber):

Okay so I'm thinking is uh... \vec{v}_f equals \vec{v}_i , one of those type of problems, um, UAM-type thing, okay [UAM stands for uniform accelerated motion]. Gravity is definitely in the mix, oh good, okay. So let's see first I'm going to draw a nice little picture cause it helps me think.

In order to segment this section of the protocol, I determined where the subject made transitions in thought. For

example, after identifying the problem as a "UAM-type thing" the subject moved to discussing gravity and then to delineating a start to the solution path. Therefore, I segmented this section as follows:

Segment one: Okay so I'm thinking is uh... V_f equals V_i , one of those type of problems, um, UAM-type thing, okay [UAM stands for uniform accelerated motion].

Segment two: Gravity is definitely in the mix, oh good, okay.

Segment three: So let's see first I'm going to draw a nice little picture cause it helps me think.

These segments were then coded using a coding scheme developed for this study. The coding scheme was based on research which showed that experts and better problem solvers use different strategies as summarized in Table IV. 6,7,9,10,12,22–28 The main segments of the coding scheme were as follows:

- (i) metacognition,
- (ii) understanding or reading the problem,
- (iii) problem analysis, and
- (iv) solution method.

The coding scheme was a bit complex with up to three additional sublevels. For example, metacognition was subdivided into planning, monitoring, and evaluating, while problem analysis was divided into problem classification, text searches, and graphing. Problem classification was further sorted into statements which identified models, problem

types, or equations. The metacognitive subcategories of planning, monitoring, and evaluating were also further divided. The entire coding scheme is located in Appendix D.

The first level determinations were the easiest ones to complete and had high inter-rater agreement. An example of an understanding or reading comment is the following from M14:

The rabbit gives chase in instant the turtle passes him as shown on the velocity time graph below. When do the two have the same velocity? [reading problem—stops at first question] So this is velocity time [points to dotted line on graph].

The subject is holding the problem and reading it while interpreting the graph that came with the problem statement. Whereas, NM18 below is completing a different type of statement:

Let's see here... Um... I realize I should put a table of contents in my notebook for days like this. Wait hold on what's this? That's not what I want either. Um... (*flipping pages*). Where's (*inaudible*) looking for... [looking through notebook but can't find what wants].

NM18 is completing a problem-analysis task specifically in this case searching her notebook for an equation. Determining the sections of the transcript that demonstrated students actually solving problems were some of the easiest to determine as can be seen in the following excerpt from NM7:

Equals negative big number 14×10^6 . Okay and then we'll subtract by the quantity 50 000 times 3000, ..., 29×10^6 . Yeah. And then we divide by 10 000. Equals -200 ... no 29 000 m/s. x component of the asteroid. [finish inputting known values and solves for final velocity]

In this instance NM7 is algebraically solving for final velocity. After a segment was coded as a problem-analysis statement I would then need to determine if the student was classifying the problem or searching text, etc. Once the statement was determined to be classifying the problem statement I then proceeded to identify if the student was identifying a model, problem type, equation, or algebraic representation. When participants were classified as identifying a model they did not have to use the name of the model since the nonmodelers would not have been trained in the specific model names. Rather, identifying a model was loosely interpreted as identifying the concept that was associated with the problem. For example, in the following excerpt M5 has finished reading a problem and says:

All right so we need to find the distance. So we need to figure out the model that we can use for mass and time to figure out the distance and it's going to be accelerating 9.8 m/s so... I always forget the constant acceleration model....

M5 is identifying that he wants to deploy the constant acceleration model. This segment would be coded as identifying a model. In the following excerpt NM2 has finished reading problem 5 dealing with types of slides and he says:

Um... Okay... So let's see... I think its energy.

While NM2 does not use the term energy model he is identifying the concept that must be deployed in order to solve this particular problem. This section of NM2's verbal protocol would be coded as identifying a model. The following excerpt from NM18's verbal protocol shows a distinct difference from the two above. NM18 had been reading problem 2 and immediately after reading the word collision in the problem he said:

Okay great. Um... I know I have a problem like this in here somewhere... Now... Um... well, let's see... [starts to look through notebook then gives up on it]....

NM18 then went in search of the similar problem type in his notebook. Therefore, he would not be coded as identifying a model but as identifying a problem type. As illustrated, identifying a problem type is recognizing that this particular problem was similar to some past problem solved by the subject via a surface feature within the problem.

The coding scheme allows one to determine if the participants were identifying equations or algebraic representations linked to a model. Participants were coded as identifying an equation if the equation selected came after the subject had previously mentioned looking for an equation, searched an equation list, or identified a problem type. For example, NM18 above indicated that he was looking for a problem type at which point he started to look through his notebook and then he said:

Okay um... (*whispering*) \vec{F}_{net} equals... I'm thinking maybe this would be it. I'm thinking maybe I want \vec{F}_{net} equals MA [writes $\vec{F}_{\text{net}} = m\vec{a}$].

In this instance this segment would be coded as identifying an equation. A student was coded as identifying an algebraic representation only if they had initially identified the model that pertained to that problem within the last two statements. After reading the problem that was designed to use the constant acceleration model M5 stated a desire to use the correct model and then immediately said:

I know it's like... Wait we don't even know... [writes $\frac{1}{2}\vec{a}(\Delta t)^2$].

Well... $\vec{x}_f = \frac{1}{2} \dots 0.5 \dots$ Um... plus zero so it's... That should be it right there. [fills in the equation with known values].

This segment would have been coded as identifying an algebraic representation as this is one of two algebraic representations contained in the constant acceleration model. In addition, this segment demonstrates that the algebraic representation seems to be chunked with the model since immediately after identifying a model students usually within one statement stated the algebraic representation attached to that model. This is very similar to the expert problem-solving strategies summarized in Table IV. There were a number of these instances observed in the modeling students' verbal protocols. In a number of cases I could not distinguish between the two possibilities and the statement was coded as could not distinguish. Due to the strictness of the coding

definition the proportion of could not distinguish was quite high and similar in number for both groups.

Any equation or algebraic representation was then coded as either forward chaining or working backward. A strict definition was adhered to in this case. A subject was considered to be working backward if an equation or algebraic representation selected contained the variable that was asked for in the problem statement or followed from the selection of a problem type. The only exception to this was when the algebraic representation directly followed the declaration of the model to be used in which case it could contain the unknown variable. For example, the excerpt above from NM18 would have been coded as working backward since the equation mentioned contains the unknown variable and it was preceded with the declaration of identifying a problem type. The excerpt from M5 would have been coded as forward chaining since it directly followed the declaration of the model to be used.

The metacognitive statements involved comments where the participants were actively regulating their cognitive processes in order to solve the problem. After the determination that a statement was a metacognitive statement I would have to decide what type of metacognitive statement it contained. The following excerpt from NM7 is a classic metacognitive statement:

Wait, no. I did something wrong here. Okay this is the force going down. Okay this is the force going down. The water stops the ball here... 5 m/s. Okay... Um...

He realizes that he did something wrong and then starts interpreting his knowledge. This would be coded as a monitoring comprehension task. M30 made the following comment after she identified that one object in the problem was moving at a constant velocity:

So they have the same velocity when the velocity of the rabbit is 5 so it's just a question of when these two lines intersect.

In this instance M30 is planning how to solve the problem by actually describing the solution path in words. The following excerpt is from M20 following her initial algebraic solution for problem 4 in the verbal protocol:

Why didn't I... see that on the graph (*laugh*). It was on the graph; it was right there... It's easy, look there's a square there and then there's two half triangles there and then you know that there's two squares there... [showing area under curve by counting squares would have been easier].

This is an example of a metacognitive statement since she is monitoring the process of solving the problem while evaluating her algebraic solution. The next step would be to determine whether the statement referred to checking the approach taken, information used, appropriateness of the equation or answer. The subject in this case already has an answer and is now checking if that answer is correct; therefore, this statement would be coded as an evaluating answer statement via an alternate path—in this case a graphical one. An example of a metacognitive statement where the subject, M14, is evaluating whether the answer makes sense as seen below:

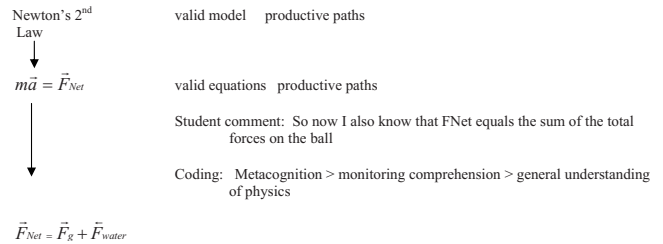


FIG. 7. Sample segment of physics protocol code.

I'm not sure if that makes sense cause it's pretty high. It's really, really high. Hmm... So this one goes from positive... negative to positive since they hit each other. And it's really big; this one's small. And has negative velocity. And impacts with the smaller one, which has a high velocity coming this way. So this goes the opposite direction (*inaudible*) but this one goes this way and it would have to be... Okay, think that... Unless I had a calculator error I think that's right.

A second rater coded a random 30% of the participants after being trained to use the verbal coding scheme. The rater was blind to condition from which each student came. These codes were checked for inter-rater reliability using Cohen's kappa correlation (a nonparametric correlation). The K values ranged from 0.95 for the first level of the code to 0.87 for the fourth level.

A physics protocol code was constructed for each student to determine the transitions in solution path, physics errors, and associated metacognitive statements produced. A metacognitive code segment of one of the physics protocols is shown in Fig. 7. The physics code was analyzed when students were on a productive solution path (a path that could lead to a correct solution) in order to determine problem-solving behaviors when they were surer of their solution paths.

C. Hypotheses

The protocols will reveal the following:

- Nonmodelers' problem-solving strategies will be more "novicelike" while modelers' will be more expertlike.
- Modelers' will produce a greater proportion of metacognitive or evaluation statements when on a productive path.
- Modelers' will catch and correct a greater percentage of physics errors.

D. Results of verbal protocol data

All of the quantitative results noted in this section include only the talk aloud protocol data and none of the retrospective interview. When the retrospective data were included in the analysis the proportions discussed below increased in value. It was predicted that the modeling students would demonstrate greater amount of expertlike problem-solving

TABLE V. Summary of statistics for problem-solving behaviors.

Statement type	Modelers average (%)	Nonmodelers average (%)	<i>p</i> value
Understand or read	21	25	$p < 0.012$
Searching text and notebook	2.2	11	$p < 0.006$
Graphical statements made	9	5	$p < 0.0005$
Classifying problem via models	33	16	$p < 0.003$
Classifying problem via problem types	0.7	8	$p < 0.005$
Identifying equations	18	45	$p < 0.001$
Identifying algebraic representations	28	9	$p < 0.0005$
Working backward toward solution	26	52	$p < 0.0005$
Working forward toward solution	70	47	$p < 0.0035$
Algebraic solutions	62	83	$p < 0.0005$
Graphical solutions	28	7	$p < 0.0005$

strategies as defined in expert novice problem-solving studies.^{6,9,22–28} It was predicted that the modeling students would identify models or principles to solve the problems while nonmodelers would focus on equations and problem surface features. It was predicted that the modelers would be more flexible in solution path since the modeling method promotes the chunking of representations around basic models of physics. To determine if the predictions were correct the verbal protocols were analyzed to determine the proportion with which each group of students identified principles instead of equations, solved problems via algebraic vs graphical methods, looked at text and equation lists more often, and solved problems via working forward vs working backward. Since modeling students are always asked to defend and explain their solution choices, it was predicted that they would use a greater proportion of metacognitive strategies such as more evaluation of their answers and methods when on a productive solution path. The physics protocol codes were used to determine the number of physics errors produced and the percentage that were caught by each group.

The study findings were quite dramatic. The proportion of coded statements produced per group and their significance level are included in Table V. As predicted the modelers identify the model associated with the problem significantly more than the nonmodelers demonstrating a reliance on deep structure (i.e., models). In addition, the nonmodelers are more likely to identify the problem type associated with the problems showing a reliance on surface features. The nonmodelers identify an equation to solve the problem 45% of the time whereas the modelers demonstrate this identification only 18% of the time. On the other hand, the modelers identify an algebraic relationship 28% of the time vs the 9% demonstrated by nonmodelers. The implications for these approaches are tremendous. By identifying problems based on the models used to solve them the modelers are demonstrating the use of deep structure and a more coherent knowledge structure. The modeling students should have an easy search space since they only have six models to distinguish between. After they select the model via a breadth search across all of their models all the representations (algebraic, graphical, diagrammatic, and verbal) chunked with the model immediately become available for the students allow-

ing them to be more efficient (i.e., should take less time to reach a solution) and demonstrate more flexibility. The nonmodelers' identification of problem types and equation lists is not as efficient a strategy. The nonmodelers' knowledge structure could consist of equations and solution methods sorted based on surface features. Thus, when selecting a solution strategy the nonmodeler would initially select a surface feature to explore such as ramps. However, numerous solution methods would be associated with the concept of ramp. Therefore, the student would have to deploy each solution method until they determine if it is possibly a correct procedure. Thus, the student would go through a number of depth searches as they work through each possible solution method. However, the modeling student selects a solution method based on models such that once the appropriate model is selected there is only a limited number of different representations that can be utilized and all representations would produce the same solution. Therefore, the modeler would spend more time doing an initial breadth structure before committing to a solution strategy. The nonmodelers produce algebraic solutions to a significantly greater extent than the modelers (83% vs 62%). The graphical solutions are even more telling with the modelers producing a significantly greater number of these solutions (28% vs 7%). Of the participants who chose to complete guess and check solutions 89% of the modelers utilized a graphical strategy vs no graphical strategy usage by the nonmodelers. This finding was statistically significant, assuming equal variances ($M > NM$, $t_{10} = 4.472$, $p < 0.0005$). The nonmodelers produced 100% algebraic guess and check solutions, while the modelers only chose this path 11% of the time. This finding was also statistically significant, assuming equal variances ($NM > M$, $t_{10} = -4.472$, $p < 0.0005$). These results highlight the greater flexibility of the modeling students in terms of solution paths open to them especially when the results of the graphical problem on the PS task from study 1 are considered. A chi-square test of solvers vs nonsolvers in study 1 demonstrated that there were no significant differences in the ability of the participants to solve graphically based problems ($p < 0.1$). Therefore, the modeling students are making better use of their abilities during problem-solving situations.

The statements coded as metacognitive were analyzed in a similar fashion. It was discovered that while both groups

produced similar proportions of metacognitive statements across the entire protocol the modeling students produced a greater proportion when on a productive solution path (27% vs 21%, $M > NM$, $t_{28} = 2.009$, $p < 0.03$). In addition, when on a productive path, the modelers had a higher proportion of evaluating or checking statements (43% vs 29%, $M > NM$, $t_{28} = 2.131$, $p < 0.02$). The modelers were evaluating their answers to a greater extent than the nonmodelers (54% vs 33%, $M > NM$, $t_{28} = 1.83$, $p < 0.043$). Most surprisingly the protocols showed that when the modelers evaluated their answers they were doing so via alternate solution paths 23% of the time while the nonmodelers never evaluated their answers in this fashion ($M > NM$, $t_{28} = 3.692$, $p < 0.001$). This result demonstrates that the modelers are much more flexible in their solution approach. The chunking of the representations with the model may allow this to occur. In addition, 3 out of 11 of the nonmodelers checked their answers by comparing them to problem types while none of the modelers checked their answer via this technique (30% vs 0%, respectively). This difference in search by problem type percentage is indeed a significant difference between the two populations ($p < 0.025$).

I also discovered that 13 out of 19 modelers (68%) check to see if the answer made sense when only 3 out of 11 nonmodelers (27%) did this. A chi-square test of the number of students checking their answers to see if they made sense was significantly different between the two groups ($p < 0.05$).

While there were not significant differences between proportions of approach evaluations if you look at what the students are doing when evaluating the approach taken there is a significant difference. When evaluating the approach taken the participants seem to do one of three tasks: they evaluate the steps in the path, they revise the approach, or they see if the approach makes sense. The only type of evaluating approach statement that proved to be significantly different between the two groups was the proportion of checking completed to determine if the approach made sense. None of the nonmodelers attempted to see if the approach made sense while 19% of the modeling statements were coded completing this task.

E. Physics errors vs correction rates

The number of physics errors and error correction rates differed significantly as summarized in Fig. 8. The modeling group on average committed 1.95 errors, while the nonmodeling group on average committed 2.91 errors ($NM > M$, $t_{28} = -2.24$, $p < 0.017$). This means that the modeling group produced only 40% of all the errors created. Not only did the modelers make fewer errors, they discovered and corrected a greater percentage of the errors. Levene's test for equality of variance determined that equal variances could not be assumed in the case of the correction rates. However, the modelers had a significantly greater correction rate of 23.1%, while, the nonmodelers correction rate was 5.3% ($M > NM$, $t_{22} = 2.09$, $p < 0.024$).

An exploration of the physics codes was conducted to determine the differences and similarities in physics errors

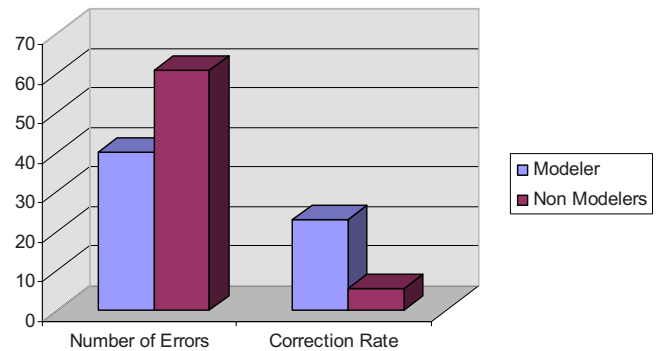


FIG. 8. (Color online) Graph of percentage of errors produced and correction rates per group. The modeling students commit significantly fewer errors while correcting a greater percentage.

produced and caught between the modeling and nonmodeling groups. There were a total of 11 types of errors produced by all participants. There are several commonalities in the types of errors. The same percentage of participants in the nonmodeling and modeling groups ignored forces on objects ($\chi^2 = 0.072$, $df = 1$, $p < 0.789$) and confused the final velocity with average velocity during solutions ($\chi^2 = 0.639$, $df = 1$, $p < 0.424$). For example, in one problem the majority of the students forgot to include the force of gravity in the calculation of the net force. In several problems, the participants would calculate the final velocity and then substitute this value in the average velocity when determining the position of an object.

Two types of errors were committed to a significantly different amount by each group. These errors included confusing a velocity time graph for a position time graph and using variables from different dimensions in the same equation. For example, in one problem students would routinely utilize the acceleration due to gravity which is in the y dimension in the same equation with a horizontal change in position thus committing the error of utilizing variables from different dimensions in the same equation. Of all nonmodelers 90% committed this error while only 70% of modelers committed it ($\chi^2 = 4.59$, $p < 0.032$, $df = 1$). In one particular problem participants routinely confused a velocity time graph for a position time graph thus reading off position data when it was actually velocity data. This confusion between velocity vs time graphs and position vs time graphs was quite notable between the two groups and statistically significant. While only 20% of all modelers committed this error, 50% of the nonmodelers committed the same error ($\chi^2 = 4.98$, $p < 0.026$, $df = 1$).

The types of physics errors that are caught are very telling. It seems that once a subject committed the error of using variables from different dimensions in the same equation there was no turning back since not a single subject corrected this error in either group. In addition, ignoring forces seems to have been an error that is difficult to correct since none of the participants did so. The subjects as a group, both nonmodeling and modeling, seemed to believe that when finding a force all they needed to do was to multiply the mass and the acceleration to determine the amount of the force in question without taking into account other forces such as a nor-

TABLE VI. Summary of correlation coefficient and R^2 values between post-tests and verbal protocol problem classification data. Categories that are statistically significant at $p < 0.05$ are shaded.

Verbal protocol problem-analysis observed behavior	Expert score	FCI score	SF score	PS task score
Identify model	0.63/39.5	0.36/13.2	-0.51/26.4	0.45/20.5
Identify problem type	-0.38/14.4	-0.29/8.3	0.54/29.6	-0.45/20.1
Identify algebraic representation	0.47/21.7	0.56/31.4	-0.54/29	0.52/27.4
Identify equation	-0.36/12.8	-0.38/14	0.41/17.1	-0.58/34
Forward chaining	0.46/21	0.56/31.3	-0.44/19.5	0.59/34.4
Working backward	-0.48/22.6	-0.64/41	0.69/47.3	-0.62/38.8

mal force or the force of gravity on the object. Students would greatly improve their performance if they simply drew force diagrams as retrospectively many of the students, especially the modelers, realized their error as soon as a detailed question was asked. One student even said that he felt he did not need to draw a force diagram on paper since he could do one in his head.

Both groups had equal success in catching minor errors such as adding masses together when they should be considered separately, using the wrong acceleration rate, and ignoring signs. Of the errors 70% confusing final velocity with the average velocity and 50% of the errors confusing velocity time graphs with position time graphs were caught by the modelers. However, the nonmodelers did not catch any of these types of errors even when the actual number of errors, as in the case of the graphical error was greater in the non-modeling group. It seems highly unlikely that this would be a random event unrelated to the use of the modeling pedagogy. The increase in the number of errors caught might be one of the factors that allow the students to have greater problem-solving success.

IV. CORRELATION BETWEEN STUDY 1 AND STUDY 2

Based on the results of study 1 and study 2 one might ask the following: Are the scores from study 1 correlated with the behavior exhibited in study 2? Cognitive structures determined in study 1 correlate significantly with the problem-solving behaviors demonstrated. The behaviors of identifying models vs problem types, identifying algebraic representations vs identifying equations, and working forward vs working backward are much more telling expertlike traits. In the initial problem analysis the subjects either chose to identify a problem type, identify a model, or identify an equation to use to solve the problem. When identifying a model the students would key into the deep structure used in the problem and state whether the concept they were dealing with was momentum or energy, for example. On the other hand, the nonmodelers seemed to identify problem types and equations more often. The nonmodelers would spend a majority of their time searching equation lists for specific formulas. When the nonmodelers identified problem types they would key into surface features such as collisions or incline planes demonstrating a high reliance on surface features. They might see an object go down a slide and immediately

look for other incline plane problems to find a match. One would suspect that a high FCI and expert score would correlate positively with identifying a model and algebraic representations, whereas, a high SF score would correlate with identifying a problem type or equation. The correlations between the study 1 task scores and problem-analysis observed behaviors are summarized in Table VI. All of the regressions are in the direction predicted while only one is not significant: FCI score vs identifying problem type. The FCI, expert, and PS task scores all were negatively correlated with identifying problem type, identifying an equation, and working backward toward a solution, whereas, the SF score was positively correlated with these behaviors. The reverse was true for identifying a model, identifying an algebraic equation, and forward chaining. All of these results are as predicted. A student's expert and SF scores account for a large proportion of the variation in these behaviors especially when identifying a model. The students are using the knowledge structures they have constructed after the completion of the year long physics course to solve the problems in this study. These findings suggest that the variations in problem-solving behaviors are accounted for to a large extent by the knowledge structure (for example, expert cognitive structure accounts for 40% of the variance in the model identification).

The data from study 2 concerning metacognitive behaviors did not show any pattern of significant correlation values with the task scores from study 1. It is surprising that no correlation was found between the metacognitive and problem-solving behaviors since prior research has shown, as has this study, that they are important skills to use in order to be a competent problem solver.

V. DISCUSSION

This study determines how modeling and nonmodeling students differ in knowledge organization, the type and use of problem-solving and metacognitive skills, and types of errors produced. The results reported in this paper demonstrate that the modeling pedagogy seems to allow for the production of more expertlike knowledge structures and that these structures correlate positively with not only quantitative and qualitative tests but also with the problem-solving strategies deployed. In addition, the metacognitive skills used by modeling students appear to be more in line with those demonstrated by expert physicists. The high correlation

between strategies employed and the students' knowledge structures and problem-solving task scores implies that modeling students may have produced a greater number of correct answers on problem-solving tests because of the production of fewer physics errors on average and the correction of a greater number of those errors. All in all, the data presented in these studies support Brewe's supposition¹⁴ that the modeling pedagogy contributes to the development of more expertlike physics students.

VI. EDUCATIONAL OR SCIENTIFIC IMPORTANCE OF THE STUDY

By understanding the mechanisms used by modeling students' instructors can more easily diagnose after interviewing students where the problem-solving trajectory can be improved thus allowing them to become more expert problem solvers. In addition, the knowledge of error difficulties can allow instructors to conduct activities that make students confront the errors that have been shown to be the most difficult for students to rectify.

Trends in International Mathematics and Science Study (TIMSS) and National Assessment of Educational Progress (NAEP) data have shown that traditional classes often fail our students.²⁹ The findings reported here allow for a cognitive perspective of why modeling instruction improves traditional learning and a possible developmental approach. If curriculum based on models can be developed for other areas then it may allow students to develop expert behaviors in a more consistent manner in these disciplines.

The protocol study also points out that the modeling students while on a trajectory from novice toward expertness still demonstrate a number of novice qualities. The findings of this study will allow for pedagogical and curricular changes to be made in the modeling instruction materials that will allow the students to continue to improve in the areas of difficulty demonstrated. For example, curricular should stress to a greater extent the traits that modelers showed themselves to be lacking such as diagrammatic usage. The students did not see the need to draw force diagrams and as one student said "they can easily be drawn in your head" implying that there was no need to draw them on paper. The modeling curriculum needs problem sets that persistently require the students to draw force diagrams in order to enjoy success and internalize the need for these types of representations. This would be an important change since flexible use of diagrams is a trait of better problem solvers and experts.^{26,30}

The modeling students experienced great difficulty sorting energy problems. The reason behind this difficulty especially for modelers should be investigated since energy is a principle which threads through all science subjects and is an important model for students to master in order to understand the connections between the science disciplines. The methods used in this study could be used to determine if an energy thread such as that used by Brewe¹⁴ allows for a different type of knowledge structure organization producing a structure more in line with the expert structures produced by physics experts of Chi *et al.*^{6,7} such that there would be only

two models or principles, energy and momentum.

This study demonstrates that the teaching of metacognitive skills may be extremely important if students are to find and correct errors. The final step of analyzing the solution must be scaffolded to a greater extent to allow more students to internalize what to look for when evaluating an answer. If the use of metacognitive skills that good problem solvers demonstrated in this study was highlighted for students and practiced by them then their ability to detect and correct errors should improve. Additional research in the area of error correction and associated metacognitive activity will allow educators to develop curriculum to help students to become expert problem solvers in all domains.

Finally, the results reported here suggest that physics education researchers should not rely solely on paper and pencil tests such as the FCI to test for the efficacy of course improvements or new courses. Researchers must be mindful that while some course changes might increase FCI scores it does not necessarily demonstrate that the students are closer to a more expertlike problem-solving behavior. It might be important for researchers to evaluate the overall cognitive changes in their students in addition to their ability to answer conceptual questions correctly.

VII. QUESTIONS FOR FUTURE RESEARCH

This study only compared modeling students to those taught in traditional high school classes. In the future comparisons should be made between modeling students and those being taught with other approaches such as inquiry oriented curriculums. In addition, it would be interesting to determine if the problem-solving strategies utilized by the students in this study are observed not only when they are solving rich complex problems but also when solving simplistic problems.

The use of the card-sort task would allow for the completion of longitudinal studies looking at the trajectory of the change in knowledge organization over time in order to answer any number of questions such as follows:

- (i) Is the change linear over the course of one school year or is there more of a quadratic relationship with a sudden rearranging of the students' knowledge?
- (ii) Is there a transition in knowledge structure from a surface feature orientation to questions-asked to expert organization?
- (iii) In addition, does the knowledge organization continue to become more expertlike even without additional classroom physics training?
- (iv) As the students move from a first- to a second-year physics course, how do their knowledge organizations continue to progress?
- (v) Will the modeling students collapse the basic models over time in order to produce a true expertlike structure based on fundamental principles? There was some evidence in the card sorts that suggest that this might be occurring since the better problem solvers had started to collapse models together such as constant velocity and constant acceleration. Would this collapse continue as students progress to-

ward becoming physics experts and at what rate would it occur?

This result demonstrates the need for more research in this area in order to produce a clearer understanding of metacognition and its connection to problem solving.

Most importantly, these results demonstrate the need for more research in the area of metacognition, problem solving, and error correction. There have not been any studies to date that considered metacognitive and problem-analysis tasks students utilize when solving a problem. Research in this area could allow educators to develop curriculum to help students to become expert problem solvers in all domains.

VIII. SUMMARY

The data demonstrate that modeling students seem to develop a knowledge structure that is not only expertlike and based on models but also correlates highly with improved performance and expert problem-solving behaviors. The data suggest that expertlike problem solving and metacognitive strategies allow for better performance due to the production of fewer errors and the correction of more errors. All in all, the qualitative and quantitative data obtained demonstrate that the modeling pedagogy is an effective method by which students can be guided along the path to becoming expert problem solvers. The data indicate that modeling instruction helps students to

- (1) identify effective solution paths,
- (2) produce fewer errors, and
- (3) correct a larger percentage of the errors they produce.

Note that the author may be contacted for a sample copy of the video transcription coding document.

ACKNOWLEDGMENTS

I would like to extend my sincere thanks to John R. Hayes, Ruth Chabay, Bruce Sherwood, Marcia Lovett, Dan Malone, and Matt Kolymeyer for their significant contributions in the development and analysis of this research project.

APPENDIX A: SUPPLEMENTARY AUXILIARY MATERIAL

See separate auxiliary material for examples of problem-solving tasks.

APPENDIX B: SUPPLEMENTARY AUXILIARY MATERIAL

PROBLEM-SOLVING TASK

See separate auxiliary material for the PS task in its entirety.

APPENDIX C: SUPPLEMENTARY AUXILIARY MATERIAL

VERBAL PROTOCOL PROBLEMS

See separate auxiliary material for the problems used in the verbal protocols.

APPENDIX D: SUPPLEMENTARY AUXILIARY MATERIAL

VERBAL PROTOCOL CODING SCHEME

See separate auxiliary material to view the entire coding scheme.

¹M. Wells, D. Hestenes, and G. Swackhamer, A modeling method for high school physics instruction, *Am. J. Phys.* **63**, 606 (1995).

²R. Hake, Interactive-engagement versus traditional methods: A six thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).

³D. Hestenes and M. Wells, A mechanics baseline test, *Phys. Teach.* **30**, 159 (1992).

⁴D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).

⁵J. Vesenka, P. Beach, G. Munoz, F. Judd, and R. Key, A comparison between traditional and “modeling” approaches to undergraduate physics instruction at two universities with implications for improving physics teacher preparation, *Phys. Teach. Educ. Online* **1**, 3 (2002) [http://www.phy.ilstu.edu/jpteo/issues/jpteo1\(1\)june02.pdf](http://www.phy.ilstu.edu/jpteo/issues/jpteo1(1)june02.pdf)

⁶M. T. H. Chi, P. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).

⁷M. T. H. Chi, R. Glaser, and E. Rees, in *Advances in the Psychology of Human Intelligence*, edited by R. J. Sternberg (Erlbaum, Hillsdale, NJ, 1982), Vol. 1.

⁸D. H. Hinsley, J. R. Hayes, and H. A. Simon, in *Cognitive Pro-*

cess in Comprehension, edited by M. Just and P. Carpenter (Erlbaum, Hillsdale, NJ, 1977).

⁹P. T. Hardiman, R. Dufresne, and J. P. Mestre, The relation between problem categorization and problem solving among novices and experts, *Mem. Cognit.* **17**, 627 (1989).

¹⁰J. Mestre, R. J. Dufresne, W. J. Gerace, P. T. Hardiman, and J. S. Touger, Promoting skilled problem solving behavior among beginning physics students, *J. Res. Sci. Teach.* **30**, 303 (1993).

¹¹R. J. Dufresne, W. J. Gerace, P. T. Hardiman, and J. P. Mestre, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, *J. Learn. Sci.* **2**, 307 (1992).

¹²A. Schoenfeld, *Mathematical Problem Solving* (Academic, Orlando, FL, 1985).

¹³Z. Mevarech and B. Kramarski, IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms, *Am. Educ. Res. J.* **34**, 365 (1997).

¹⁴E. Brewster, Inclusion of the energy thread in the introductory physics curriculum: An example of long-term conceptual and thematic coherence, Ph.D. thesis, Arizona State University, 2002.

¹⁵R. L. Keith, Correlation between the consistent use of a general problem solving strategy and the organization of physics knowledge, Ph.D. thesis, University of Minnesota, 1993.

- ¹⁶E. A. Silver, Recall of mathematical problem information: Solving related problems, *J. Res. Math. Educ.* **12**, 54 (1981).
- ¹⁷J. Schinka, I. Weiner, and W. Velicer, *Handbook of Psychology, Research Methods in Psychology* (Wiley, New Jersey, 2003), Vol. 2.
- ¹⁸R. Thordike, *Measurement and Evaluation in Psychology and Education* (Prentice-Hall, New Jersey, 1997).
- ¹⁹A. Hirai, The relationship between listening and reading rates of Japanese EFL learners, *Mod. Lang. J.* **83**, 367 (1999).
- ²⁰Relative difference was calculated by subtracting the lowest result from the highest result dividing by the average of the two results and then multiplying by 100.
- ²¹Normalized gain $\langle g \rangle = (\%post - \%pre) / (100 - \%pre)$.
- ²²J. L. Snyder, An investigation of the knowledge structures of experts, intermediates and novices in physics, *Int. J. Sci. Educ.* **22**, 979 (2000).
- ²³A. Dhillon, Individual differences within problem-solving strategies used in Physics, *Sci. Educ.* **82**, 379 (1998).
- ²⁴M. Hegarty, R. E. Mayer, and C. A. Monk, Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers, *J. Educ. Psychol.* **87**, 18 (1995).
- ²⁵J. H. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980b).
- ²⁶A. B. Lewis, Training students to represent arithmetic word problems, *J. Educ. Psychol.* **81**, 521 (1989).
- ²⁷J. Larkin and F. Reif, Understanding and teaching problem-solving in Physics, *Eur. J. Sci. Educ.* **1**, 191 (1979).
- ²⁸D. P. Simon and H. A. Simon, in *Children's Thinking: What Develops?*, edited by R. Siegler (Erlbaum, Hillsdale, NJ, 1978).
- ²⁹W. Schmidt, C. McKnight, and S. Raizen, *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education* (Kluwer, Dordrecht, The Netherlands, 1997).
- ³⁰J. H. Larkin and H. A. Simon, Why a diagram is (sometimes) worth ten thousand words, *Cogn. Sci.* **11**, 65 (1987).