Modeling instruction in mechanics

Ibrahim Abou Halloun* and David Hestenes
Department of Physics, Arizona State University, Tempe, Arizona 85287
**Now at the Lebanese University*

(Received 1 August 1984; accepted for publication 9 July 1986)

Modeling theory was used in the design of a method to teach problem solving in introductory mechanics. A pedagogical experiment to evaluate the effectiveness of the method found positive results.

1. INTRODUCTION

This article reports on the first attempt to improve physics teaching by applying the modeling theory of instruction set forth in Ref. 1. The theory was applied to the design of a specific instructional method, and the effectiveness of the method was assessed by a pedagogical experiment. The experiment can therefore be regarded as a partial test of the instructional theory. A full scale test of the theory is out of the question at this time, because that would require a wholesale revamping of available instructional materials. Moreover, the basic principles of the theory are not sufficiently specific to be tested directly. Rather, they determine the direction of a research program for the development of instructional design. The purpose of our pedagogical experiment, therefore, was to ascertain whether our first steps in that direction actually improve instruction.

For good reasons that need not be discussed here, problem solving performance is usually taken as the principle measure of student understanding in a physics course. The poor performance typical of most students on physics examinations suggests that conventional methods for teaching problem solving are far from optimal. For this reason we decided to apply the instructional theory directly to the design of instruction in problem solving. We did that in two ways. First we employed the conventional lecture method for a systematic exposition of modeling principles and techniques in solving mechanics problems. As we did not expect that approach to be very effective, at least by itself, we developed another approach which makes greater use of the instructional theory.

We turn now to consider what went into the design of our teaching method before describing our pedagogical experiment and discussing its results and implications.

II. THE METHOD OF PARADIGM PROBLEMS

From previous research, we know that the initial "common sense" (CS) knowledge of students has a significant detrimental effect on their performance in physics courses, and conventional instruction is ineffective in correcting defects in this knowledge (Ref. 1, 2, 3, and references cited therein). A taxonomy for analyzing defective *CS factual knowledge* about mechanics has been developed[2]. Defective *procedural knowledge is* evident in the "formula-centered" approach to problem solving employed by most beginning students. This information should be used to improve instructional design. According to our instructional theory[1], we need an instructional method which promotes a *model-centered* approach to problem solving, and the method should have a dialectical component to promote the substitution of "Newtonian knowledge" for defective CS knowledge.

To meet these specifications, we have devised what we call the *method of paradigm problems.* The method has three major features:

(a) systematic design and selection of paradigm problems for intensive study;

(b) a dialectical teaching strategy;

(c) a gradual introduction of modeling theory and techniques.

Let us elaborate on each of these features in turn.

Our method is based on a careful selection of a small set of paradigm problems in the scientific domain of interest, specifically, particle mechanics in our pedagogical experiment. We say "paradigm problems" because the problems are chosen to be characteristic of problems in the entire domain under study, and their solutions are to be used as models for the solution of other problems in the domain. Paradigm problems are designed and selected using principles of our instructional theory. Two general kinds of consideration are involved. First, we seek problems which evoke and challenge the typical *common sense* misconceptions about the subject domain. Second, we select a set of problems which require for their solutions the *complete* system of modeling techniques which we aim to teach, including domain specific modeling techniques as well as general modeling strategy and tactics. No problem will require all the techniques for its solution, but each problem will require a nontrivial application of at least one important technique.

It is important to keep the number of paradigm problems to a minimum so there will be time to study each problem thoroughly. It is equally important to aim for a set of paradigm problems which is *complete* in the sense that it confronts *all* typical misconceptions and embraces *all* the important modeling techniques. A set of five to ten paradigm problems seems to be sufficient for teaching Newtonian particle mechanics at the introductory level. We do not present a particular set here, because we have not settled on a set which we think is optimal. However, Reif and Heller[4,5] give several good candidates supported by the kind of analysis which is needed before they can be used effectively as paradigm problems.

The manner of engaging students in the study of paradigm problems is critical, if the paradigm problems are to perform their intended instructional function. Students need time to recognize and equilibrate discrepancies between Common Sense and Newtonian concepts as well as to understand the nature and purpose of specific modeling techniques. Theory suggests that a situation like this calls for a dialectical teaching strategy[1]. Accordingly, an *interactive* problem-solving approach like the following is appropriate. The instructor begins the attack on a given problem by soliciting from the students on what information to select from the givens and how to represent and use this information. Specifically,

1. The instructor asks students to list all relevant information.
2. Students furnish ideas which the instructor writes on a blackboard (or overhead projector).
3. A group discussion follows to separate relevant from irrelevant information.
4. The instructor asks students what to do with the selected information.
5. Students furnish problem solving tactics and strategies which the instructor outlines on the board. Student suggestions can usually be classified into two or three main types.
6. The instructor asks each group of students to defend its tactics/strategies. Different groups are asked to criticize tactics/strategies offered by other groups. Students are asked to support their arguments with appeals to facts and laws of physics.
7. If the conflict between groups is resolved and an appropriate method of solution emerges, the instructor summarizes the major steps that need to be implemented and proceeds to solve the problem.

8. Otherwise, the instructor introduces new ideas for students to consider until sufficient information for solving the problem is available.

As a rule, at least 20 minutes is needed for the analysis and solution of a paradigm problem by this interactive approach.

In leading the class study of paradigm problems, the instructor is guided by the principles and techniques of modeling theory which are introduced as they are needed. These ideas are summarized and integrated in a system of handouts which are distributed to the class gradually as the instructor decides that the students are ready to appreciate and use them. For Newtonian mechanics we recommend the following system of handouts:

I. An outline of the general model development strategy (such as Fig. 1 in Ref. 1).

II. A series of handouts summarizing information from the knowledge base to be used in the first two stages of model development; specifically, handouts on: (a) motion description, (b) interaction description, (c) motion laws, and (d) interaction laws (see Tables 1-5 in Ref. 4 for good examples) .

III. Handouts on model ramification (in particular, Fig. 2, of Ref. 1).

Each of the handouts is worthy of some discussion when it is distributed, since it summarizes an important body of information. After a handout has been distributed, students are encouraged to refer to it in problem solving practice and in subsequent discussions of paradigm problems whenever it is relevant. After the complete set of handouts has been distributed, students should be able to justify any step in problem solving by reference to an appropriate handout. Moreover, they should master the information on the handouts so completely that the handouts will not be needed for examinations. Note that the handouts will not be used in problem solving in the same order that they are distributed. For example, Fig. 2[1] will be one of the first handouts distributed, since it is concerned with kinematics which is studied before dynamical concepts are introduced.

Of course, students must practice to become skillful problem solvers. And for efficient learning, they must practice good technique. The method of paradigm problems provides exemplars and an organized system of principles to direct the attention of students toward what they should practice. Ideally, the class discussion of a paradigm problem should be immediately followed by student practice on problems specifically selected to reinforce the lessons learned in the discussion. Unfortunately, we fell short of this ideal in our pedagogical experiment, because an adequate methodological analysis and classification of all the assigned problems was too big a task for us to complete.

III. DESCRIPTION OF THE EXPERIMENT

The experiment was designed to test the following two predictions:

I. Student achievement in physics can be improved by incorporating a systematic discussion of modeling techniques into class lectures.

II. Student achievement can be further improved by employing the method of paradigm problems in recitation classes.

All subjects in the experiments were students in the first semester of University Physics at Arizona State University in the fall of 1983. Characteristics of the student population, the course content, and the textbook were the same as for the courses in Ref. 3. In addition, the courses were taught in a conventional lecture-recitation format as described in Ref. 3. The students were divided into four groups, a *control group* (CG) of 119 subjects, and three *treatment groups* (TG1, TG2, TG3) consisting of 235, 64, and 20 subjects, respectively.

Lectures to the treatment groups were delivered by one of us (DH) in two separate classes. One class consisted of TG1 subjects only, while the remaining subjects attended the other class. Lectures to both classes were based on a modeling approach to mechanics along the lines described in Ref. 1. The distinctive ingredient of the lectures with greatest relevance to our experiment was a detailed discussion of the descriptive stage in developing particle models, illustrated by specific applications to problem solving. Particular attention was paid to the representation of information by diagrammatic techniques such as maps, graphs, and force diagrams, as well as the translation of information into mathematical form. Moreover, the rationale for each technique was carefully explained. The explicit discussions of modeling concepts and procedures continued through mid semester until the sections on single particle mechanics were completed. Thereafter, beginning with many particle systems, the course was taught along conventional lines, with only incidental references to modeling. This was done to see if we could detect any residual benefits from the initial instruction in modeling.

The University Physics course includes a weekly 50-minute recitation class of no more than 25 students. The three treatment groups differed in the recitation instruction they received. Recitations for TG1 were conducted by experienced graduate teaching assistants who were unaware of the nature and objectives of the experiment. Recitations of TG2 and TG3 were taught by one of us (IH) using the method of Paradigm Problems, although the handouts distributed differed somewhat from those recommended in our discussion of the method. A single hour per week is insufficient to take full advantage of the method, so TG3 was given an additional two hour recitation session each week for the first seven weeks of the semester.

Since the students in TG3 were *required* to attend extra recitation sessions, they had to be volunteers. The 20 students in TG3 were selected randomly from a group of 73 volunteers. Subsequent performance of the volunteers who were not chosen did not deviate significantly from that of the TG1 and TG2 groups in which they remained. So initial motivation was not an important factor distinguishing TG3 from the other groups.

The control group attended lecture and recitation classes which were completely independent of the treatment classes. The control group was given conventional physics instruction by a professor and teaching assistants who were unaware of the nature and objectives of the experiment. However, the control and treatment classes used the same textbooks, had the same daily schedule of topics to be covered and the same set of recommended homework problems.

For baseline data, the competence of students in all groups was measured by the validated physics and mathematics pretests described in Ref. 3. Students were classified into three competence classes:

(a) *High competence* for a total physics-math pretest score of 40 or more (out of 79 possible);

(b) *Average competence* for a total score between 30 and 40; or

(c) *Low competence* for a total score of 30 or less.

Student achievement was measured by pre-post-test gain on the mechanics diagnostic test of Ref. 3 and by performance on four course exams, including a final exam and three “one-hour exams” given at approximately four week intervals during the semester. The course exams were composed entirely of physics problems. To minimize any possibility of “teaching the tests”, the problems were selected mainly from exams used by previous instructors without participation of the TG2 and TG3 instructor (IH). To make direct comparisons of achievement possible, the exams were the same for all treatment and control groups, except for differences on the third exam, and a uniform grading scheme was imposed.

Table I

Analysis of variance on the various groups' competence as measured by the diagnostic tests

	Math pretest		Mechanics pretest		Percentage in competence subgroups	
Group	**Mean**	**S.D.**	**Mean**	**S.D.**	**Average**	**Low**
CG	17.45	6.37	17.90	6.26	23%	34%
TG1	14.08	6.42	17.21	6.42	34%	47%
TG2	14.43	6.24	18.03	6.28	35%	36%
TG3	14.17	6.16	16.89	6.00	20%	55%
All TG's	14.15	5.87	17.49	6.32	33%	45%
	$F = 22.76$	$p = 0.0001$	$F = 0.25$	$p = 0.61$		

Highest possible score was 33 on the math pretest, and 36 on the mechanics pretest.
In tables I-VI the value for F is obtained by an ANOVA test (Ref. 8) .
p is the probability that variations in column values may have occurred by chance (Ref. 8).

IV. RESULTS AND CONCLUSIONS FROM THE EXPERIMENT

All treatment groups achieved a higher gain than the control group on the mechanics diagnostic test and performed better on examinations. The diagnostic tests enabled us to compare these groups with previous groups under other instructors. Table I gives the average scores for all groups on the physics and mathematics diagnostic pretests. Differences in the mechanics pretest scores are insignificant, but the control group shows a higher mathematics competence. Comparison with the results in Ref. 3 shows that these scores are typical of introductory physics classes at ASU.

The mechanics diagnostic test was repeated at the end of the semester. The pre-post-test gains for all groups and their average- and low-competence subgroups are given in Table II. Comparison with the results of Ref. 3 shows that the control group gain is comparable to gains of classes taught by other instructors, but the gains of all treatment groups exceed the gains of any previous class, including one taught by the same professor (DH). The greater gain of TG1 is significant, but only the gain of TG3 is impressive by comparison. What do these results tell us?

In Refs. 2 and 3, we established that a student's pre-posttest gain on the mechanics diagnostic test can be regarded as a measure of the transition from common sense concepts to Newtonian concepts of motion. In the lectures to the treatment group, no special efforts were made to deal with common sense misconceptions, so the greater gains of TG1 over CG are reasonably attributed to alterations of common sense concepts induced by instruction in modeling, though the evidence for this conclusion is admittedly weak. On the other hand, a systematic effort to eliminate common sense misconceptions were made in TG2 and TG3 recitations, and Table III shows clear evidence for its success. The gain was greatest for low-competence students. This result is perhaps more impressive when we note that the average mechanics post-test score for the low-competence students in TG3 were nearly the same as that of students who received an A grade in the course. This is clear evidence that the difference between high- and low-competence groups is not simply a difference in innate abilities. Instruction can be effective! However, the absolute posttest scores are still far from satisfactory, so the instruction leaves a lot to be desired.

Table II

Analysis of variance on the various groups' physics pretest-posttest gain

	All participating S's				All S's		Average		Low	
	Pretest[b]		Post-test		Gain		Subgroup Gain		Subgroup Gain	
Group	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
CG	19.10[c]	6.26[a]	22.92	6.57	3.82[c]	4.43	4.82	3.75	3.81	4.24
TG1	17.96	6.58	23.87	6.29	5.91	4.74	5.16	4.50	8.59	8.45
TG2	18.54	6.22	25.52	5.76	6.98	5.41	5.91	4.35	10.37	7.17
TG3	17.47	6.16	27.06	4.68	9.58	5.51	8.50	10.60	11.56	5.64
F =	0.83		2.15		7.32		0.42		3.07	
p =	0.48		0.09		0.0001		0.74		0.03	
All TG's:										
	17.84	6.30	24.26	6.81	6.42	6.23	5.27	4.59	8.90	7.97
F =	0.25[a]		1.47		11.17		0.17		7.89	
p =	0.61[a]		0.23		0.0009		0.68		0.006	

[a] Values for comparing CG with the total TG.

[b] This data is only for students who completed the course, so the pretest data differs somewhat from that in Table 1.

[c] This CG data was also reported in Table I of Ref. 3 as results for professor D.

Table III

Analysis of variance on the various group's primary achievement in PHY 115

	First exam		Second Exam		Primary Index[b]	
Group	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
CG	53.28[a]	23.46	25.11	19.43	40.13	18.53
TG1	58.70	23.45	35.69	20.71	50.09	18.27
TG2	61.28	25.23	41.18	25.20	53.82	22.84
TG3	66.45	13.35	48.55	10.30	58.11	9.11
F =	2.95		11.97		10.28	
p =	0.03		0.0001		0.0001	
All TG's	59.73	23.37	37.77	21.41	52.41	18.89
F =	6.45		26.54		26.52	
p =	0.01		0.0001		0.0001	

[a] Test scores are given as percentages of a perfect score in Tables III–VI.

[b] The primary index is the average of scores on the first two exams, computed only for students who took both exams.

Turning now to an analysis of performance on course exams, our first concern is to establish that the instruction given to the control group is typical of conventional instruction. The lectures were given by a respected colleague; they followed closely the form and content of a conventional textbook, and example problems on the various topics were worked out regularly. We have already seen that the diagnostic test results for the control group were typical of previous classes. For additional external evidence, the first l-h exam was identical to one designed and given previously by another professor. The average score of the control group on this exam were 53% (S.D. = 23.46), which is not significantly different from the score of 51% (S.D. = 19.89) for the previous class. For these reasons, we regard the performance of the control group as typical of class performance under conventional instruction.

Average scores on the first two 1-h exams for all groups are given in Table III. The first exam covered mainly elementary kinematics and particle motion under a constant force. The second exam was concerned more generally with particle dynamics. These two exams covered the entire (half semester) period of special instruction for the treatment groups. The average difference between treatment and control groups rose from 6.45% on the first exam to 12.66% on the second exam, suggesting a cumulative effect of the special instruction. As a performance measure for the entire period of special instruction we use the mean of the first two exam scores, which we call the *index of primary* achievement.

To establish that the performance difference between treatment and control groups was, in fact, a consequence of the special instruction, we analyzed individual responses on the second exam. The treatment groups had been given exceptionally thorough instruction in systematically identifying all forces on a body, correctly representing their properties on force diagrams and translating this information into equations of motion. Our analysis confirmed that this accounts for the performance difference. The exam consisted of four problems for which solutions would be facilitated by drawing force diagrams. In fact, diagrams were almost a necessity to solve two of the problems. About 85% of the trained subjects, but only 40% of the controls, tried to draw force diagrams for these two problems. Fifty percent of the trained and 20% of the controls were able to draw correct diagrams. Of those students in both groups, about 75% were able to determine correct equations of motion. Nearly all students who drew wrong diagrams were unable to determine the correct equations. And less than 10% of the students who did not draw diagrams were able to determine correct equations. Most of those who found the correct equations of motion were able to solve the problem completely; the failures, of course, were due mainly to mathematical deficiencies.

The real effect of the special training may be most accurately indicated by the comparative fractions of students who drew correct diagrams (50% *vs.* 20%). In other words, the real effect is more than twice the 12.5% difference in average test scores. This agrees with our estimate that perhaps 50% of the TG1 and TG2 groups ignored the special training for various reasons, such as the fact that there is no mention of it in their textbook and many students attend large lecture classes irregularly. The better performance of TG3 is explained in part by much better compliance with the special training. When we take into account the fact that the TG3 has a much larger percentage of low competence students then the other groups (Table IV), we see that the primary achievement of TG3 was clearly superior to that of all other groups. Indeed, the primary achievement of the TG3 low-competence students was better than that of high-competence students in the control group CG. Note that this is consistent with our results on diagnostic test gains.

The analysis of exam results supports our belief that the superior achievement of TG3 was due principally to the nature of the special instruction and is not merely a practice effect from the extra hours of recitation. Unfortunately, for practical reasons we were unable to check this by setting up a control group with the same number of extra hours of conventional recitation. However, our previous experience with extra hours of *conventional* problem solving recitations had convinced us that it has little effect on student achievement.

Table IV

Analysis of variance on average and low competence students' primary achievement[a] in the various groups.

	Average competence			Low competence		
Group	**S's (%)**	**Mean[a]**	**S.D.**	**S's (%)**	**Mean**	**S.D.**
CG	33%	33.20	13.19	28%	31.30	16.90
TG1	37%	52.89	17.52	41%	41.04	15.39
TG2	35%	59.65	20.65	37%	43.33	20.79
TG3	20%	59.00	4.36	55%	57.50	10.05
		F = 12.53	p = 0.0001	F = 7.00	p = 0.0003	
All TG's	35%	54.49	17.99	41%	43.19	15.46
		F = 35.09	p = 0.0001	F= 8.21	p= 0.004	

[a] The *index of primary achievement* is the mean of the first two exam scores.

Table V

Analysis of variance on the various groups' performance in system of particle mechanics (Third exam)

	All subjects		Average competence			Low competence		
Group	**Mean**	**S.D.**	**S's (%)**	**Mean**	**S.D.**	**S's (%)**	**Mean**	**S.D.**
CG	35.49	23.34	30%	25.88	16.98	28%	26.09	18.68
TG1	44.10	20.64	38%	45.02	19.67	39%	35.96	18.33
TG2	49.20	17.97	38%	48.50	14.06	39%	39.60	12.04
TG3	48.17	13.46	17%	47.50	4.94	61%	44.30	1.34
	F= 5.56	p= 0.001	F= 7.21	p = 0.0003		F= 2.51	p = 0.05	
All TG's	45.43	19.70	36%	45.51	18.38	41%	87.94	16.99
	F= 14.24	p = 0.0002		F= 21.09	p = 0.0001		F= 6.13	p = 0.01

Our conclusion that improved performance is a consequence of our special training is supported by results of Heller and Reif[5], which we learned about after our experiment was completed. In an exceptionally well-controlled experiment, they validated a prescriptive model for generating useful descriptions of scientific problems. Our special training emphasized nearly all the important elements of their model. Consequently, our results confirm the pedagogical value of their model in a real classroom situation. Their excellent formulation and explication of scientific description should be widely used in physics teaching.

Results in Table V of the third exam indicate that earlier special instruction in modeling transferred to better achievement on new material. This exam included problems on rigid body dynamics and particle collisions. The control group was given a different exam on the *same* topics, but experienced instructors agreed that every problem on it was easier than the corresponding problem on the exam given to the treatment groups. Nevertheless, the comparatively better performance of the treatment groups was maintained at all levels of competence: (a) about 10% for high-competence students, (b) 20% for average competence, and (c) 12% for low competence. For high- and average-competence students, the differences of the three treatment groups on this exam were insignificant. But the better performance of low-competence students in TG3 continued. This was in spite of the fact that many failing students in TG1 and TG2 withdrew from the course after the second exam, but no low-competence students withdrew from TG3.

Table VI shows that the comparatively better performance of the treatment groups persisted through the final exam, which covered new material on thermodynamics and emphasized material in the last half of the semester. The semester index reported in the table is based on weekly quiz grades as well as all course exams. Different quizzes were given to treatment and control groups but the scores were normalized to the same average value in computing the index.

An overall comparison of achievement for the various groups and subgroups is given in Table VII. Grade cutoffs were nearly the same for treatment and control groups. Overall, the fraction of students achieving a grade of C or better in the treatment groups were 15% larger than in the control group.

Table Vl

Analysis of variance on the various groups' semester achievement

	All subjects		Average competence			Low competence		
Group	**Average**	**S.D.**	**S's %**	**Average**	**S.D.**	**S's %**	**Average**	**S.D.**
CG	44.39[a]	16.81	30%	36.12	10.62	27%	36.19	13.50
	45.45	17.13	30%	36.20	10.63	27%	38.48	14.11
TG1	51.57	14.60	37%	53.12	12.83	39%	46.64	14.33
	53.04	13.96	37%	54.36	11.68	39%	46.62	12.43
TG2	55.59	15.96	38%	54.71	12.50	39%	48.76	16.87
	56.95	16.36	38%	57.74	13.29	39%	48.81	16.14
TG3	55.22	9.88	17%	52.33	2.83	61%	52.30	9.17
	56.93	8.14	17%	53.10	5.66	61%	55.31	9.39
	F = 7.11	p = 0.0002		F = 11.95	p = 0.0001		F = 4.28	p = 0.007
	F = 7.92	p = 0.0001		F = 13.91	p = 0.0001		F = 4.41	p = 0.006
All	52.68	14.61	36%	53.30	12.52	41%	47.03	14.32
TG's	54.15	14.16	36%	55.03	11.84	41%	47.00	12.66
	F = 18.35	p = 0.0001		F = 35.83	p= 0.0001		F = 9.49	p = 0.003
	F = 20.72	p = 0.0001		F = 41.21	p= 0.0001		F = 7.32	p = 0.009

[a] Results for each group are shown for the final exam in the upper line, and for the semester index in the lower line. The semester index is an average over all quizzes and exams, including the final.

Table Vll
Final grade distribution in the various groups

	Group	Percentage of students with a specific grade					
		A	B	C	D	E	W
All subjects	CG	8	9	23	17	15	28
	TG1	8	15	30	9	5	33
	TG2	15	13	24	11	3	34
	TG3	5	40	40	5	0	10
	All TG's	9	17	29	9	4	32
Average competence	CG	0	3	24	21	24	27
	TG1	9	17	38	7	3	26
	TG2	21	11	32	10	0	26
	TG3	0	0	75	0	0	25
	All TG's	11	15	37	8	2	26
Low competence	CG	0	6	17	18	17	42
	TG1	0	9	28	11	8	44
	TG2	0	9	25	15	9	42
	TG3	0	45	45	9	0	0
	All TG's	0	11	29	11	7	42

Surely the most notable result in the entire experiment was the achievement of the TG3 low-competence students shown in Table VII. None of these 11 students withdrew, and only one failed to achieve a C or B grade. By contrast, the dismal performance of the CG low-competence students shown in the table is typical of conventional instruction; for we established in Ref. 3, using statistics from several previous classes, that about 80% of low-competence students failed to complete the course with a grade of C or better. Thus, the TG3 instruction overcame tremendously unfavorable odds. We have additional evidence that this kind of success can be replicated. Evidently, many students with low-initial competence can indeed be salvaged!

V. DISCUSSION AND RECOMMENDATIONS FOR RESEARCH

To be frank, we were surprised by the results of our pedagogical experiment. To begin with, we were surprised to find such clear evidence for improved student performance as a result of mere lectures on modeling theory and techniques. This has given us new respect for the lecture method. The best previous evidence known to us indicated that in large physics classes, wide variations in the apparent quality of lectures, from doggedly competent to charismatic, produce negligible variations in student performance on exams. But our results now indicate that the performance of *some* students can be improved by variations in the *content* of the lectures bearing on the technical foundations for performance. This gives us some hope that further refinement in lectures on the technical aspects of modeling may have more beneficial effects.

Nevertheless, we remain cognizant of severe inherent limitations in the lecture method. The lecture method is a *didactical* (or *preceptive)* method, best suited to setting forth a particular line of thought to an audience which shares the preconceptions of the lecturer. Therefore, it is more appropriate for advanced courses in physics than for introductory courses where the majority of students are not yet accustomed to the modes of scientific thinking. For teaching introductory courses, a *dialectical,* (or *evaluative)* method is needed to provoke students to analyze and evaluate their own preconceptions, to examine the grounds for common sense and scientific beliefs. This point has been eloquently made by Arnold Arons[6] on many occasions. And the results of our experiment give it some empirical support. Of course, some combination of didactical and dialectical methods is needed at every level of instruction.

We were most surprised by the large diagnostic test gains of low-competence students who received dialectical instruction in modeling, and by high correlations of these gains with exam performance. Surprised, because the results conform to our instructional theory so perfectly! Yet all previous research indicated that such large gains should be extremely difficult to achieve, because common sense preconceptions are so stable. More recently, Minstrell[7] has reported impressive success in teaching the Newtonian force concept with an intensive dialectical method. This supports our theoretical prediction that defective common sense knowledge can be corrected by a suitable dialectical teaching method. It gives us good reason to believe that major improvements in instruction can be achieved in this direction.

A variation of our pedagogical experiment carried out a year later gave some additional information. One of the teaching assistants in the control group of our original experiment became familiar with our instructional theory and method. He decided to follow a *model-centered strategy* for problem solving in teaching recitations, but he was encouraged to do things his own way without rigidly following the format used in our experiment. The result was a dramatic improvement in his teaching. This was obvious in an unprecedented wave of enthusiasm among the students, who raved about the helpfulness of his recitation sessions. He became the most popular teaching assistant ever in our department. He, himself, attributed his success mainly to the modeling approach, which helped him identify key factors in problem solving and develop a new order and coherence in his teaching. An objective measure of his success was obtained by comparing the performance of his students with those of a control group in the same way as in our experiment. The overall relative gains with respect the control were about the same as in our experiment, with the average competence students doing better but the low competence students not as well. However, on the diagnostic test his students did no better than the control group. We believe this means that his success in teaching procedural knowledge may have been greater than in our experiment, but he failed to deal adequately with the factual misconceptions which interfere most with the learning of low competence students. He agrees that he made no special effort to expose such misconceptions in this teaching. By the way, Figs. 1, 2, and 3 of Ref. 1 were used as handouts in this experiment. Figure 2 proved to be very helpful both in class and in student self-study. But Figs. 1 and 3 were not emphasized in recitations, so most students ignored them. Thus, it remains to be seen whether they can play a valuable practical role as handouts in support of modeling instruction.

It should be evident that effective implementation of the method of paradigm problems requires considerable skill and preparation on the part of the instructor. Mastery of the subject matter is insufficient though it is necessary. As we have described it, the method is best suited for classes of moderate size, say between 10 and 40 students. In larger classes, reciprocal interaction between students and instructor is more difficult to achieve, so it might be better to use a variant of the method in which the instructor proposes conflicting alternatives to be evaluated by discussion (instead of soliciting alternative from the students). Variants of the method might also be implemented on interactive computer systems. Surely, extensive research and development would be required to perfect the method for a given course.

To improve problem solving instruction in mechanics, we think the greatest need is for a systematic analysis and classification of available textbook problems so they can be employed more efficiently in instruction. There are many excellent problems in the textbooks, but they are organized only by subject, rather than any specific pedagogical objectives they might subserve. References 1 and 2 suggest that a pedagogical classification should answer such questions as "What common sense misconceptions might be probed or challenged by this problem?" and "What specific modeling techniques are needed for the solution?". To develop such a pedagogical classification would be a formidable task. But besides its immediate value in instruction, it would become an important tool in further research.

1. D. Hestenes. 1987. "Toward a Modeling Theory of Physics Instruction". *Am. J. Phys.*, **55**, 440-454.
2. I. A. Halloun and D. Hestenes. 1985. "Common Sense Concepts about Motion". *Am. J. Phys.*, **53**, 1056-1065.
3. I. A. Halloun and D. Hestenes. 1985. "The Initial Knowledge State of College Physics Students". *Am. J. Phys.*, **53**, 1043-1055.
4. F. Reif and J. I. Heller. 1982. "Knowledge Structure and Problem Solving in Physics". *Educational Psychologist*, **17**, 102-127.
5. J. I. Heller and F. Reif. 1984. "Prescribing Effective Human Problem Solving Processes: Problem Description in Physics". *Cognition and Instruction*, **1**, 177-216.
6. A. Arons. 1981. "Thinking, Reasoning and Understanding in Introductory Physics Courses". *Phys. Teach.*, **19**, 166-172.
7. J. Minstrell. 1984. "Teaching for the Development of Understanding of Ideas: Forces on Moving Objects". In the *1984 Yearbook of the Association for the Education of Teachers in Science*. ERIC, Ohio State University, Columbus, Ohio.
8. G. Anderson *et al.*. 1975. *Encyclopedia of Educational Evaluation* . Jossey-Bass, London.