

The initial knowledge state of college physics students

Ibrahim Abou Halloun^{a)} and David Hestenes

Department of Physics, Arizona State University, Tempe, Arizona 85287

An instrument to assess the basic knowledge state of students taking a first course in physics has been designed and validated. Measurements with the instrument show that the student's initial qualitative, common sense beliefs about motion and causes has a large effect on performance in physics, but conventional instruction induces only a small change in those beliefs.

I. INTRODUCTION

Each student entering a first course in physics possesses a system of beliefs and intuitions about physical phenomena derived from extensive personal experience. This system functions as a *common sense theory* of the physical world which the student uses to interpret his experience, including what he uses and hears in the physics course. Surely it must be the major determinant of what the student learns in the course. Yet conventional physics instruction fails almost completely to take this into account. We suggest that this instructional failure is largely responsible for the legendary incomprehensibility of introductory physics.

The influence of common sense beliefs on physics instruction cannot be determined without careful research. Such research has barely gotten started in recent years, but significant implications for instruction are already apparent. Research on common sense beliefs about motion 1-5 has lead to the following general conclusions.

(1) Common sense beliefs about motion are generally incompatible with Newtonian theory. Consequently, there is a tendency for students to systematically misinterpret material in introductory physics courses.

(2) Common sense beliefs are very stable, and conventional physics instruction does little to change them.

Previous research into common sense beliefs has focused on isolated concepts. Here we aim for a broader perspective. This article discusses the design and validation of an instrument for assessing the knowledge state of beginning physics students, including mathematical knowledge as well as beliefs about physical phenomena. Measurements with the instrument give firm quantitative support for the general conclusions above. The instrument can be used for instructional purposes as well as further research. In particular, we recommend the instrument for use:

(1) *As a placement exam.* The instrument reliably identifies students who are likely to have difficulty with a conventional physics course, so these students can be singled out for special advisement or instruction.

(2) *To evaluate instruction.* The instrument reliably evaluates the general effectiveness of instruction in modifying a student's initial common sense misconceptions.

(3) *As a diagnostic test* for identifying and classifying specific misconceptions. This will be discussed in a subsequent paper.

II. ASSESSMENT OF A STUDENT'S BASIC KNOWLEDGE STATE

To evaluate physics instruction objectively, we need an instrument to assess a student's knowledge state before and after instruction. In the following sections we discuss the design and validation of such an instrument. The instrument consists of two tests: (a) a *physics diagnostic* test to assess the student's qualitative conceptions of common physical phenomena and (b) a *mathematics diagnostic* test to assess the student's mathematical skills. Both tests are intended for use as *pretests* to assess the student's initial knowledge state. The mechanics test is also intended for use as a *post-test* to measure the effect of instruction independent of course examinations.

A. Design of the physics diagnostic test

The first course in physics is concerned mainly with mechanics, and mechanics is an essential prerequisite for most of the rest of physics. Therefore, the student's initial knowledge of mechanics is most critical to his course performance, so we can restrict our attention to that domain of physics. Now, it would be far from sufficient simply to test a student's initial knowledge of Newtonian mechanics. Rather, we need to ascertain the student's common sense knowledge of mechanics, for it is the discrepancy between his common sense concepts and the Newtonian concepts which best describes what the student needs to learn. As Mark Twain once observed, "It's not what you don't know that hurts you. It's what you know that ain't so!"

Newtonian theory enables us to identify the basic elements in conceptualizations of motion. On one hand, we have the basic *kinematical concepts* of position, distance, motion, time, velocity, and acceleration. On the other hand, we have the basic *dynamical concepts* of inertia, force, resistance, vacuum, and gravity. We take a student's understanding of these basic concepts as the defining characteristics of his basic knowledge of mechanics. Our list of dynamical concepts may look a bit strange, to a physicist, but the particular items on the list were chosen to bring to light major differences between common sense and Newtonian concepts. We refer to a knowledge state derived from personal experience with little formal instruction in physics as a "common sense knowledge state." As a rule, it differs markedly from the "Newtonian knowledge state" of a trained physicist.

To assess the student's basic knowledge of mechanics, we devised the *physics (mechanics) diagnostic test* presented in the Appendix. The test questions were initially selected to assess the student's qualitative conceptions of motion and its causes, and to identify common misconceptions which had been noted by previous investigators. Various versions of the test were administered over a period of three years to more than 1000 students in college level, introductory physics courses. Early versions required written answers. Answers reflecting the most common misconceptions were selected as alternative answers in the final multiple-choice version presented in the Appendix. In this way we obtained an easily graded test which can identify a spectrum of common sense misconceptions.

A student's score on the diagnostic test is a measure of his *qualitative understanding* of mechanics. We shall see that statistically it is quite a good measure because of its reliability and predictive validity. We believe also that it is a theoretically sound measure, because the diagnostic test is concerned exclusively with a systematic assessment of basic concepts. One could not

expect satisfactory results from the typical "physics achievement test" which tests for knowledge of isolated physical facts.

B. Validity and reliability of the mechanics test

The face and content validity of the mechanics test was established in four different ways. First, early versions of the test were examined by a number of physics professors and graduate students, and their suggestions were incorporated into the final version. Second, the test was administered to 11 graduate students, and it was determined that they all agreed on the correct answer to each question. Third, interviews of 22 introductory physics students who had taken the test showed that they understood the questions and the allowed alternative answers. Fourth, the answers of 31 students who received A grades in University Physics were carefully scrutinized for evidence of common misunderstanding which might be attributed to the formulation of the questions. None was found.

The reliability of the mechanics test was established by interviewing a sample of students who had taken the test and by a statistical analysis of test results. During the interviews, the students repeated the answers they had given on the written test virtually without exception. Moreover, they were not easily swayed from their answers when individual questions were discussed, and they were usually able to give reasons for their choices. It seemed clear to the interviewer that the students' answers reflected stable beliefs rather than tentative, random, or flippant responses. This impression is strongly confirmed by the high reproducibility of responses on retests.

To compare test score distributions for different (but comparable) groups tested at different times, the Kuder-Richardson Test⁶ was used. The values obtained for the KR reliability coefficient were 0.86 for pretest use, and 0.89 for post-test use. These unusually high values are indicative of highly reliable tests. A similar comparison of score distributions for written answer and multiple-choice versions of the tests gave comparable results, confirming the conclusion that the multiple-choice version measures the same thing as the written version, but more efficiently.

The possibility of relevant test-retest effects was eliminated by two procedures. First, the post-test results of one group of 29 students who had not taken the pretest was compared with those of a larger group in the same class who had taken the pretest. The means and standard deviations for both groups were nearly identical. Second, a group of 15 students was given the post-test shortly after midterm and again at the end of the semester. The mean test score and standard deviation for this group were, respectively, 22.79 and 3.60 for the first post-test, and 23.58 and 3.26 for the second. This tiny change in score shows that most of the improvement between pretest and post-test scores which we discuss later occurs in the first half of the semester, as one would expect.

C. Mathematics diagnostic test

Our mathematics diagnostic test was designed to assess specific mathematical skills known to be important in introductory physics. The final version consisted of 33 questions, including (a) ten algebra and arithmetic items, (b) eight

trigonometry and geometry items, (c) four items on graphs, (d) six reasoning items, and (e) five calculus items.

To get a multiple-choice test which is as valid as a written test, our first version of the test required written answers from which we selected the most common and significant errors as alternatives to the correct answer on the multiple choice version. It is worth mentioning that the errors are not completely random; rather they tend to fall in patterns indicating common misconceptions. As Piaget noted more than half a century ago, the errors can tell us a lot about how students think. Unfortunately, instructors still pay scant attention to errors in the mistaken belief that it is pedagogically sufficient to concentrate on correct answers. We will not analyze mathematical misconceptions, but we will be concerned with a parallel analysis of physical misconceptions in a subsequent paper.

To maximize the predictive validity of the mathematics test, we began with a long list of questions, and for the final version of the test we selected only questions which correlated significantly with achievement in physics. The resulting test was judged by experienced physics instructors to be rather difficult for beginning students. The point is that the ability to do the easier math problems is hardly sufficient for success in physics.

Table I. Average diagnostic test results by course and professor. Maximum Scores: 36, for the physics diagnostic test; 33, for the mathematics diagnostic test including five calculus items which were omitted in College Physics.

Professor	Number of S's	Math Pretest Mean (s.d.)	Physics		
			Pretest	Post-test	Gain
University Physics					
A	97	17.25 (5.37) [52%]	18.47 (5.29) [51 %]	23.23 (4.94) [65%]	4.76 [13%]
B	192	16.80 (6.21) [51%]	18.39 (5.14) [51%]	23.13 (4.81) [64%]	4.74 [13%]
C	70	19.56 (5.81) [59%]	18.06 (5.95) [50%]	22.91 (5.81) [64%]	4.85 [13%]
D	119	17.45 (6.37) [53%]	19.10 (6.26) [53%]	22.92 (6.57) [64%]	3.82 [11%]
College Physics					
E	82	10.48 (4.58) [37%]	13.48 (5.00) [37%]	19.00 (5.16) [53%]	5.52 [15%]
E	196	10.19 (4.51) [36%]	13.33 (5.09) [37%]	Not Available	
F	127	9.75 (4.38) [35%]	14.43 (5.16) [40%]	Not Available	
High School Physics					
G	24 (honors)		10.96 (3.28) [30%]	18.88 (5.02) [52%]	7.92 [22%]
G	25 (general)		10.83 (3.85) [30%]	15.80 (4.34) [44%]	4.97 [14%]

A KR reliability coefficient of 0.86 for our mathematics I test shows that its reliability is comparable to that of the mechanics test.

A copy of the math test is not included in this article, since we will not be concerned with specific questions in it, and such tests are fairly easy to construct. We shall, however, evaluate the predictive power of the test.

III. RESULTS AND IMPLICATIONS FOR TEACHING

The math and physics diagnostic tests have been used to assess the basic knowledge of nearly 1500 students taking University or College Physics at Arizona State University, and of 80 students beginning physics at a nearby high-school. ASU is a state university of about 40,000 students located metropolitan Phoenix area, which has a population of about 1½ million. ASU will accept highschool graduates in the upper half of their class, and any student transferring from community colleges with passing grades. The local community colleges will accept any high school graduate. Thus our results may be expected to be typical of a large American urban university with open enrollment. University Physics at ASU is a two-semester, calculus based introductory physics course, but we will be concerned here with the *first semester only*. At ASU, about 80% of the students in this course are declared engineering majors. Although calculus is a corequisite rather than a prerequisite for University Physics at ASU, nearly 80% of the beginning students have already completed one or more semesters of college calculus. The first semester of University Physics is concerned mainly with mechanics, including some fluid mechanics, as well as elementary kinetic theory and thermodynamics.

College Physics at ASU covers nearly the same subject matter as University Physics, but without using calculus. Trigonometry is a prerequisite for the course. Most of the students take the course because it is required for their majors.

Table I presents diagnostic test results for classes in University Physics taught by four different professors, and for classes in College Physics taught by two different professors. Considering the nature of the diagnostic test in the Appendix, the average scores on the tests appear to be very low. Interpretation of these results will be our main concern, but for comparison we first take note of the test results for high school students.

We were surprised by the extremely low mechanics pretest scores of the high school students shown in Table I. Their average is only a little above the chance level score of 7.3 on the multiple-choice test. All scores were less than 20, except for one student with the score of 28, who incidentally dropped out of school before completing the physics course. The honors students were selected for high academic performance or achievement test scores, but their physics intuitions are evidently no better than anyone else's. Note that the post-test score of the high school honor students is within the range of pretest scores for the college students in University Physics. However, the post-test score of high school students in General Physics is about two points higher than the pretest scores for students in College Physics. This difference seems to be explained by the fact that about 55% of the students in College Physics had not taken physics before, although those who had averaged only two points better on the physics pretest. At any rate, diagnostic test scores of high school physics students

should be investigated further to make sure that the low pretest scores are typical. If they are, then they provide clear documented evidence that physics instruction in high school should have a different emphasis than it has in college. The initial knowledge state is even more critical to the success of high school instruction. The low scores indicate that students are prone to misinterpreting almost everything they see and hear in the physics class.

A. Prediction of student performance in physics

To what degree does a student's performance in physics depend on his initial knowledge state? A measure of this dependence is obtained by correlating course performance with scores on the math and mechanics pretests and other initial data. A statistical analysis of these correlations leads to the following general conclusions.

- (1) Pretest scores are *consistent* across different student populations.
- (2) Mechanics and mathematics pretests assess *independent* components of a student's initial knowledge state.
- (3) The two pretests have higher *predictive validity* for student course performance than all other documented variables combined.

Course grade is a measure of course performance. In all of the courses discussed here, the student's course grade was determined almost entirely by performance on examinations consisting primarily of physics problems. Thus the student's course grade and total exam score are measures primarily of physics problem solving performance.

As we have already noted, the consistency of diagnostic test scores is indicative of test reliability. The high consistency across different class populations is obvious from Table I, without any fancy statistical analysis. The consistent difference of nearly 1 s.d. between scores for University and College Physics classes is indicative of the different science and math backgrounds, as well as academic orientations of the two populations. The fact that scores on the mechanics pretest improve with instruction for all classes is another indication of consistency. A finer statistical analysis shows that the differences between groups are random. There was not a single question on which students performed consistently better or worse from one group to another.

Taken at face value, the mechanics and math tests appear to assess different kinds of knowledge. The former is concerned with physical intuition while the latter is concerned mainly with mathematical skills. It is true that a physicist's intuitions about motion have mathematical counterparts. But the same cannot be said about the common sense intuitions of students. Therefore, we should expect little correlation between the two test scores within the student population. This has been confirmed by statistical analysis, in particular, by low values for correlation coefficients. For an early version of the two pretests, we obtained a correlation coefficient of 0.32. Further analysis revealed a correlation of 0.34 between scores on certain reasoning items in the math test and scores on the mechanics test. When these items were omitted, the correlation between math and mechanics pretests dropped to 0.19.

The distribution of student pretest scores according to course grades in University Physics is given in Table II; a significant correlation between pretest score and grade is evident. The correlation between mechanics pretest scores

and total course exam scores was evaluated for three different classes in University Physics (taught by different professors). A correlation coefficient of about 0.56 ($p = 0.0001$) was found in each case, with no significant difference between classes.

Similar evaluations of the correlation between the math pretest and course performance consistently gave values for the correlation coefficient of about 0.48 for University Physics and 0.43 for College Physics. This is slightly higher than the values of 0.35-0.42 found by Hudson,⁷ perhaps because of our procedure for constructing the math pretest. These results show conclusively that the initial knowledge measured by the two pretests has a significant effect on course performance.

The predictive validity of both pretests coupled with the extremely low correlation between them, tells us that high mathematical competence is not sufficient for high performance in physics. Evidently, this explains the common phenomenon of the student who is struggling in physics even while he is breezing through calculus.

To ascertain the relative influence of other variables on course performance, we documented individual differences with respect to gender, age, academic major, and background courses in science and mathematics. Differences in gender, age, academic major, and high school mathematics showed no effect on physics performance.

High school physics background showed some correlation with performance in College Physics but none in University Physics. About 17% of the students in University Physics had previously completed College Physics, and 20% in both courses were repeating the course after a previous withdrawal. These students did no better than those who were taking the course for the first time. The combined effects of all college and math background courses, including calculus, accounted for no more than 15% of the variance. This agrees with the findings of other investigators^{3,8} that the differences in academic background have small effects on performance in introductory physics.

To assess the combined and relative effects of the diagnostic pretests and background courses in physics and mathematics, we determined the variance loading of each variable by measuring R square in a stepwise regression analysis. The stepwise variance loading is presented in Table III. Note that the combined effect of differences in student academic background accounts for only about 15% of the variance in both College and University Physics, much less than the variance accounted for by either diagnostic test alone. The two diagnostic pretests together accounted for about 42% of the variance. We are not aware of any other science or math pretest with such a high correlation. Presumably, the remainder of the variance depends mainly on the motivation and effort of the students, as well as the quality of instruction.

The R -square values in Table III provide a standard statistical measure for the predictive validity of the diagnostic tests. However, we found that better predictions can be made using student pretest scores directly. Using a linear regression analysis of course performance scores predicted by pretest scores and cutoffs established by course instructors, we predicted the grades for a University Physics class with the results shown in Table IV. For this class, 53% of the grades were correctly predicted. A higher percentage of grades were

correctly predicted for summer school courses, presumably because the initial knowledge

Table III. Variance loading on course achievement of diagnostic pretests and background courses measured by R square in a stepwise regression analysis. In a stepwise regression analysis, R square is the product correlation function for two variables after the residual effect of other variables has been removed (Ref. 6).

Variable	R Square	
	University Physics	College Physics
Physics pretest	0.30	0.32
Math pretest	0.26	0.22
Physics and math pretests	0.40	0.42
Physics courses	0.07	0.12
Math courses	0.10	0.04
Physics and math courses	0.15	0.16
All pretests and courses	0.49	0.51

The main value of the above exercises in statistical analysis is the background it provides for interpreting the diagnostic test scores. As a practical measure of the students' knowledge state, we recommend a *Competence Index* (CI) defined in terms of the combined physics diagnostic score (PHY) and math diagnostic score (MAT). We define the competence index by

$$CI = PHY + MAT$$

for University Physics, and

$$CI = 1.5(PHY) + MAT$$

for College Physics. The weight factor of 1.5 in the latter equation reflects the greater loading of the physics pretest (see Table III).

When the combined diagnostic tests are to be used as a placement exam, we recommend a classification of students into three *competence levels*:

- (a) *High*, when $CI > 40$ (max $CI = 69$);
- (b) *Average*, when $30 < CI < 40$;
- (c) *Low*, when $CI < 30$.

With probabilities greater than 0.60 in the large student population we have studied, high competence students were likely to receive an A or B course grade, average competence students were likely to receive a C grade, and low competence students were likely to receive a D or E grade. More specifically, we have found that, of the low competence students, 95% get grades of low C or less, and only 5% do better. Moreover, about 40% of the students taking physics at our university fall in the low competence class.

Clearly, low competence students can be expected to have great difficulties with college physics. Thus one of the best uses of the competence index is to identify low competence students for the purpose of special instruction or placement in a prephysics preparatory course.

It should be remembered that the CI is not a measure of intelligence. Rather, it is a measure of the difference between common sense and scientific

conceptual frameworks. The lower the CI, the greater the difficulty in understanding scientific discourse and the greater the need for sensitive student-centered instruction.

Table IV. Distribution of actual versus predicted grades.

Predicted Actual	A	B	C	D	E	Total
A	41%	50	9	00	00	
B	3	65	32	00	00	
C	00	24	70	5	00	
D	00	23	68	9	00	
E	00	00	40	60	00	
W	00	12	81	7	00	
Total						53
(A, B)	71					
(C, D)				76		
(C, D, E)				79		
Pass/Fail				90		

B. Evaluation of physics instruction

The mechanics diagnostic test can be used as an instrument to evaluate the effectiveness of instruction in improving the students' basic knowledge. Of course, instruction may have many worthwhile objectives not measured by the diagnostic test. But improvement of basic knowledge as we defined it above should be the primary objective, since such knowledge is the foundation for the whole conceptual edifice of physics.

The gain in basic knowledge as measured by the mechanics diagnostic test is given in Table I for several different University and College Physics courses. The small values (14%) for the gain indicate that conventional instruction has little effect on the student's basic knowledge state. For the courses in Table I, values of the correlation coefficient for pretest-post-test scores range between 0.60 and 0.76. These high values are statistical indicators of little change in basic knowledge.

To interpret the gain data in Table I we need to know something about the content of the courses and how they were conducted. All of the courses were conducted in a lecture-recitation format with 3 or 4 h of lecture and 1 h of recitation each week. The lectures were given by a professor to classes ranging in size from about 80-230 students. Recitation classes of 25 students or less were conducted by graduate teaching assistants. They were devoted to problem solving. The courses did not include laboratory work, but most students took introductory physics lab courses in parallel with the lectures.

The content of the courses in Table I is fairly standard. A review of the most widely used textbooks for college level introductory physics reveals that they cover certain standard topics at a fairly standard level of mathematical sophistication, and they include a large number of standard-type problems. Thus these textbooks specify a certain *standard content* for introductory physics. We refer to instruction on this standard content using the lecture-recitation format described above as *conventional physics instruction* because it is so common in

American universities. The instruction in all the courses under consideration was conventional in this sense.

Within the format of conventional instruction, wide variations in instructional style are possible. The styles of the four lecturers in University Physics listed in Table I differ considerably. Professor A is a theoretical physicist; his lectures emphasize the conceptual structure of physics, with careful definitions and orderly logical arguments. The other professors are experimental physicists, but with quite different specialties. Professor B incorporates many demonstrations in his lectures, and he expends great time and energy preparing them; he strives especially to help students develop physical intuition. Professor C emphasizes problem solving, and he teaches by example, solving one problem after another in his lectures. Professor D is an experimental physicist teaching introductory physics for the first time; he followed the book closely in his lectures. All four professors are known as good teachers according to informal peer opinion and formal evaluations by students. Indeed, Professor B has twice received awards for outstanding teaching.

Now, Table I shows that the basic knowledge gain is the same for all four of the classes in University Physics. All four classes used the same textbook (Tipler⁹), and covered the same chapters in it. Considering the wide differences in the teaching styles of the four professors, we conclude that the *basic knowledge gain under conventional instruction is essentially independent of the professor*. This is consistent with the common observation among physics instructors that the most strenuous efforts to improve instruction hardly seem to have any effect on general student performance.

The small gain in basic knowledge under conventional instruction is all the more disturbing when one considers the uniformly low levels of the initial knowledge states shown in Table I. This means that throughout the course the students are operating with a seriously defective conceptual vocabulary, which implies that they continually misunderstand the material presented. The small gains in basic knowledge "explain" the high predictive validity of the mechanics pretest; the student's ability to process information in the course depends mainly on his initial knowledge state and hardly improves throughout the course. The high predictive validity of the pretest is not intrinsic to the test; rather, it indicates a failure of conventional instruction. The more effective the instruction is in altering the basic knowledge state, the lower the predictive validity of the pretest.

The mechanics post-test score correlates more highly with course performance than the pretest score, as it should if improvements in basic knowledge improve performance. However, the post-test scores in Table I are unacceptably low considering the elementary nature of the test. Even for the A students (about 10% of the students who complete the course) the average post-test score is only about 75%. Whereas, we think that one should not be satisfied with any instruction which fails to bring all students who pass the course above the 75% level. Conventional instruction is far from meeting this standard.

IV. CONCLUSIONS

Our diagnostic test results show that a student's initial knowledge has a large effect on his performance in physics, but conventional instruction produces comparatively small improvements in his basic knowledge. The implications of failure on the part of conventional instruction could hardly be more serious, for

we are not talking about a few isolated facts that students failed to pick up. One's basic physical knowledge provides the conceptual vocabulary one uses to understand physical phenomena. A low score on the physics diagnostic test does not mean simply that basic concepts of Newtonian mechanics are missing; it means that alternative misconceptions about mechanics are firmly in place. If such misconceptions are not corrected early in the course, the student will not only fail to understand much of the material, but worse, he is likely to dress up his misconceptions in scientific jargon, giving the false impression that he has learned something about science.

The individual instructor can hardly be blamed for the failure of conventional instruction. The instructor cannot take common sense misconceptions into account without knowing what they are and how they can be changed. To be sure, every experienced instructor has acquired a store of incidental insights into student misconceptions. But this by itself leads to incidental improvements of instruction at best, and the hardwon insights of one instructor are passed on to others only haphazardly. The full value of such insights can be realized only when they are incorporated into a program of systematic pedagogical research aimed at the development of a practical instructional theory.

We submit that the primary objective of introductory physics instruction should be to facilitate a transformation in the student's mode of thinking from his initial common sense knowledge state to the final Newtonian knowledge state of a physicist. One should hardly expect instruction which fails to take initial common sense knowledge into account to be more effective than a method for integrating differential equations which ignores initial conditions.

Diagnostic test scores provide a general index of knowledge states, but for instructional purposes we need a classification of initial states that identifies specific misconceptions that need to be corrected. We will address that problem in a subsequent paper.

^{a)}Now at The Lebanese University II.

¹ A. Caramazza, M. McCloskey, and B. Green, 'Naive beliefs in sophisticated' subjects: Misconceptions about trajectories of objects, *Cognition* **9**, 117 (1981).

² A. B. Champagne, L. E. Klopfer, and I. H. Anderson, "Factors influencing the learning of classical mechanics," *Am. J. Phys.* **48**, 1074 (1980).

³A. B. Champagne and L. E. Klopfer, "A causal model of students' achievement in a college physics course," *J. Res. Sci. Teach.* **19**, 299 (1982).

⁴M. McCloskey, A. Caramazza, and B. Green, "Curvilinear motion in the absence of external forces," *Science* **210**, 1139 (1980).

⁵J. Clement, "Students' preconceptions in introductory mechanics," *Am. J. Phys.* **50**, 66 (1982).

⁶G. Anderson *et al.*, *Encyclopedia of Educational Evaluation* (Jossey-Bass, London, 1975).

⁷H. T. Hudson and D. Liberman, "The combined effect of mathematics " skills and formal operational reasoning on student performance in the general physics course," *Am. J. Phys.* **50**, 1117 (1982).

⁸W. Wollman and F. Lawrenz, "Identifying potential 'dropouts' from I college physics classes," *J. Res. Sci. Teach.* **21**, 385 (1984).

⁹P. A. Tipler, *Physics* (Worth, New York, 1982).