A Mechanics Baseline Test

By David Hestenes and Malcolm Wells

We have designed a test to assess student understanding of the most basic concepts in mechanics. The test is universal in the sense that it is limited to concepts that should be addressed in introductory physics at any level from high school through Harvard University. We have extensive data on postinstruction scores across the whole range of levels. This provides baseline data for evaluating and comparing the effectiveness of instruction at all levels. For this reason we refer to the test as the *Mechanics Baseline* (or just the Baseline).

A copy of the Baseline test is provided in the Appendix to be used in any way the instructor sees fit. We believe, however, that the best use of the test is for postinstruction evaluation, except for advanced university courses where it may be useful as a preinstruction placement exam. Of course the self-defeating practice of "teaching to the test" should be avoided, but an examination of the test could help some teachers see where their instruction can be improved. The design of the test and some of its instructional implications are discussed in the next section. The paper concludes with a discussion of the baseline data.

Test Design and Interpretation

The Mechanics Baseline test should be compared with the *Force Concept Inventory* in the preceding paper.[1] The Baseline is the next step above the Inventory in mechanics understanding. Questions on the Inventory were designed to be meaningful to students without formal training in mechanics and to elicit their preconceptions about the subject. In contrast, the Baseline emphasizes

Table 1. Newtonian Concepts on the Mechanics Baseline Test. Each concept is involved in the corresponding question. A parenthesis means that other components are significantly involved in the question.

	Question
A. Kinematics	
Linear motion	
constant acceleration	1,(2),(3)
average acceleration	(18),(23)
average velocity	25
integrated displacement	24
Curvilinear motion	
tangential acceleration	4
normal acceleration	5,(8),(12)
$a = v^2/r$	(9),(12)
B. General Principles	
First Law	(2)
Second Law	(3),8,(9),(12),(18)
dependence on mass	17,21
Third Law	(12),(13),14
Superpostition Principle	7,(5),(13),19
Work-Energy	20
Energy Conservation	10,11
Impulse-Momentum	16,22
Momentum conservation	15
C. Specific Forces	
Gravitational free-fall	6,26
Friction	(9)

Table II. Scores on the Mechanics Baseline.

Question No. †	AZ Reg. (%)	AZ Hon. (%)	AZ AP (%)	MW Reg (%)	MW Hon (%)	AVH (%)	HU Reg (%)	HU Hon (%)	HU Time seconds (Std. Dev.)
1 (B)	54	69	69	61	73	79	78	75	91(52)
2(D)	40	51	56	39	70	78	78	82	64(43)
3(E)	29	44	59	50	70	60	93	90	59(45)
4(C)	85	80	84	94	90	86	67	69	61(73)
*5(A)	1	1	3	11	40	72	18	12	50(45)
6(C)	45	44	56	61	73	53	87	96	29(24)
7(C)	8	8	25	22	40	46	36	38	80(56)
8(D)	23	30	31	72	83	67	81	92	76(51)
*9(A)	21	23	25	17	47	40	68	86	243(134)
10(E)	35	43	28	61	97	50	89	93	51(49)
11(E)	25	26	34	17	40	47	85	85	193(134)
*12(C)	12	17	9	6	17	29	24	30	147(115)
13(B)	31	37	47	56	83	69	79	82	94(76)
14(B)	50	56	75	83	93	76	87	100	41(29)
15(E)	48	47	41	56	83	79	83	90	104(88)
16(A)	16	17	9	22	47	38	60	73	39(32)
17(D)	26	33	31	22	63	60	81	81	70(50)
*18(B)	15	19	25	28	20	40	32	51	148(104)
19(C)	16	17	34	39	47	29	78	84	63(33)
20(C)	25	24	9	28	70	28	46	49	128(96)
									35(40)
21(A)	62	71	53	61	83	93	89	97	36(80)
22(B)	56	49	53	61	40	67	32	48	78(56)
23(D)	28	41	44	39	53	74	84	85	127(88)
24(A)	29	50	44	17	70	35	59	74	77(59)
25(A)	25	37	38	33	67	26	61	70	85(48)
26(E)	13	20	28	28	57	31	53	71	
Test Avg. (Std. Dev.)	32 (11)	37 (15)	39 (15)	42 (16)	62 (17)	61 (18)	66 (14)	73 (11)	38.4 min (13.2min)
Calculation	31	33	30	31	45	51	54	64	
Diagram	14	17	24	27	43	45	46	53	
Kinematics	30	39	41	39	58	57	62	68	
Number of Students	600	116	32	18	30	58	183	73	256

Notes:

This tables shows the percentage of students who chose the correct answers for each question.
AVH (PHY 105 at Arizona State University) and HU (Harvard University) are the only university classes in this table. The others are high school classes.
'Calculation is the average score on those questions (9, 11, 12, 18, 20, 21, and 22) which require the most calculation.
'Diagram' is the average score on those questions (5, 7, 12, 13, 18, 19, and 26) for which force diagrams would facilitate the solution
'Kinematics is the average score on the kinematics questions (1, 2, 3, 4, 5, 8, 9, 12, 18, 23, 24, and 25) in Table I.
†Letters in parentheses indicate correct answers
*Questions that are in more than one of the Calculation, Diagram, Kinematics categories. Questions 12 and 18 are in all three categories.

concepts that cannot be grasped without formal knowledge about mechanics. The two tests are complementary probes for understanding of the most basic Newtonian concepts. Together they give a fairly complete profile of this understanding.

Table I outlines the Newtonian concepts involved in the Baseline test along with the questions in which they appear. It will be noted that the coverage of basic concepts is quite systematic, although the coverage of Newton's first and third laws is deliberately thin because these concepts are adequately assessed by the Inventory. Thus, the Inventory and the Baseline are complementary tests in a practical sense

For the most part, the Baseline looks like a conventional quantitative, problem-solving test, though its main intent is to assess qualitative understanding. The multiple-choice distractors in the Baseline are not commonsense alternatives as they are in the Inventory, though they include typical student mistakes, which are more often due to deficient understanding than to carelessness. We excluded problems that can be solved by a simple "plug-in" of numbers into a formula.

Judged by the low scores of students at all levels, the Baseline is not an easy test. A few of the questions were extracted from Advanced Placement exams, though we found very few AP questions suited to our purpose. Less than a third of the questions require algebraic manipulation or more than one-step reasoning, and advanced concepts such as angular momentum are excluded. Student difficulties with the test appear to stem from real deficiencies in understanding the basic concepts. We aimed for a balanced coverage of these basic concepts, but we made a point of including topics that we know pose the greatest difficulty. Two deserve special mention: kinematics and conservation laws.

We think that kinematics is the most difficult topic in elementary mechanics. It may be the most fundamental as well, for, as Newton asserted in the preface to his *Principia,* it is from the motions of objects that we discover the forces. Even so, kinematics instruction is usually far from adequate, though there are well-documented techniques for doing much better.[2] As Table I shows, the Baseline test gives kinematics the attention it deserves, and the results in Table II document the general weakness of kinematics instruction. The results of questions 4 and 5 are especially significant, for they reveal widespread deficiencies in the qualitative understanding of acceleration.

To be more specific about the concepts in question, recall that one of the most comprehensive results in kinematics is the general "acceleration theorem," which asserts that at any point on an arbitrary particle trajectory, the acceleration, **a**, can be decomposed into tangential and normal components, as expressed by

$$\mathbf{a} \ = a_v \, \mathbf{e}_v + \ a_n \, \mathbf{e}_n \tag{1}$$

where the tangential component measures the rate of change of speed v

$$a_v \ = \ dv \, / \, dt \tag{2}$$

and the normal component measures the rate of change of the velocity direction $\mathbf{e}_v$ which is related to the speed by

$$a_n = v^2/r \tag{3}$$

where r is the radius of curvature of the trajectory.

      This theorem is too advanced for most high-school students, but physics teachers should understand it perfectly. Unfortunately, many do not. Indeed, in a study of expert and novice understanding of the acceleration concept, Reif and Allen[3] found that only one out of five "experts" exhibited perfect understanding, while one other exhibited quite marked deficiencies. These "experts" were all professors of physics who had recently taught introductory physics at a major university. The lesson to be learned here is not Richard Feynman's dictum that "science is the belief in the ignorance of experts," but rather that the inadequacies of kinematics instruction are far reaching indeed!

      Introductory physics should aim at least for a qualitative understanding of Eq. (1), specifically that the tangential component measures a change in the magnitude of the velocity, whereas the normal component measures a change in the direction of the velocity, so the normal component always points to the side toward which the trajectory is bending. It is this qualitative understanding that is tested on the Baseline, and it is just this that the experts failed to correctly apply in simple physics problems. Such difficulty can be attributed, in part at least, to an overemphasis on the left side of $\mathbf{f} = m\mathbf{a}$ in mechanics. Most of the professors struggled with an analysis of the forces involved in $\mathbf{f}$ when a simple examination of $\mathbf{a}$ would tell them what they needed to know. Properly taught, a qualitative understanding of acceleration in not beyond the reach of high-school students.

      Questions 4 and 5 on the Baseline probe this qualitative understanding. Most of the Harvard students who missed question 5 chose option E (see the test), most likely correctly noting that the tangential acceleration is zero or that the vertical component of motion reverses its direction at the point in question. The only quantitative application of the acceleration theorem on the Baseline test is the application of Eq. (3) to circular motion, where r is the radius of the circle. This is an appropriate level of generality for high-school physics.

      Reit[4,3] presents a careful analysis of what is required to understand acceleration, but teaching the concept effectively remains a very challenging task.
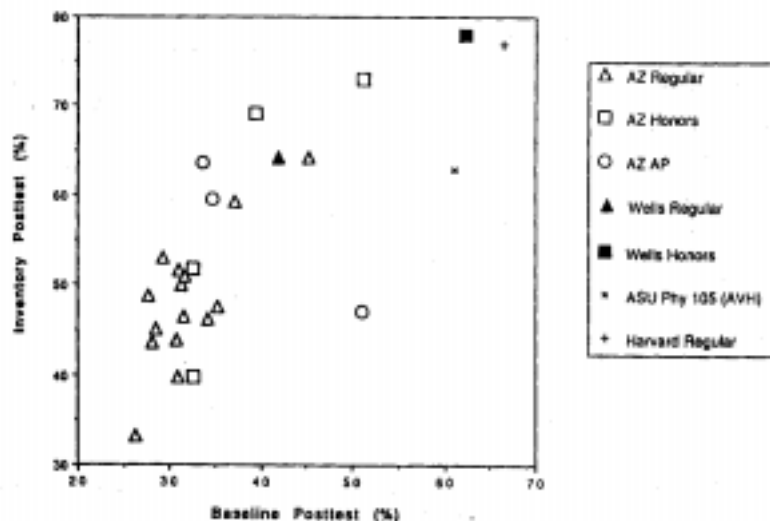

Fig. 1. Inventory vs Baseline class averages for high-school and university classes.

Concerning the conservation laws for energy and momentum, it should be noted that a full understanding involves knowing when to use them in their work-energy or impulse-momentum forms. This is probed in questions 20 and 22 on the Baseline—questions that present difficulties even to advanced students. These questions were inspired by research of McDermott and Lawson,[5] to which the reader is referred for a thorough treatment of the instructional issues.

In summary, the Baseline tests the application of Newtonian concepts to simple kinematics and dynamics of a single particle. If these topics have not been mastered, of what value is instruction on more advanced topics in mechanics? The data we have suggest: not much!

Baseline Data

Figure 1 plots inventory vs Baseline average scores for instructors listed in Tables III and IV of the preceding paper.[1] As suggested in that paper, the most noteworthy result in the figure is the placement of Wells Honors above Arizona State University and near Harvard University. In comparison with the other high-school scores, this indicates that there is great room for improvement in high-school results. Figure 2 is especially informative. It plots post test inventory vs Baseline scores for all students in the Harvard (Regular) calculus-based physics course. The inventory post test was administered about midterm, shortly after instruction and an exam on Newtonian particle mechanics. The Baseline was administered at the end of the semester, before students studied for final exams.

The calculated correlation coefficient of 0.68 between Harvard inventory and Baseline scores is fairly strong. However, Fig. 2 shows us more. Note first the clustering of points near the diagonal from the origin to the upper right-hand corner, and second that large deviations from the diagonal occur above but not below the diagonal. This supports the view that *a good score on the Inventory is a necessary but not a sufficient condition for a good score on the Baseline* or on other problem-solving tests on mechanics. We expect a good score to be necessary, because we believe that the Inventory measures the student's grasp of the Newtonian force concept. We expect it to be insufficient, because additional knowledge is required for effective problem solving.
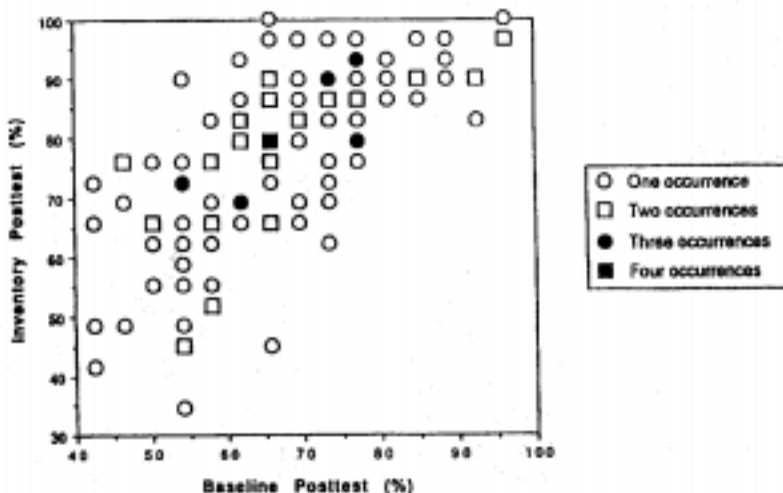

Fig. 2. Inventory vs Baseline for Harvard Regular students.

This conclusion can be refined. We suggest that a score of 60% on the Inventory is a kind of *conceptual threshold* for problem-solving competence on students below 60% on the Inventory to surpass 60% on the Baseline. This threshold effect can also explain the uniformly low Baseline scores of the Arizona physics. Below this threshold the student's grasp of Newtonian concepts is too limited for effective problem solving. Indeed, Fig. 2 shows that it is unlikely for high-school majority and the comparatively high scores of Wells Honors. In the Arizona majority the student Inventory scores are below threshold, whereas in the Wells case they are well above threshold. Indeed, our more detailed data show that more than half the students in Wells Honors scored 60% or above on the Baseline, and this is nearly double the number of students who did as well in Arizona Honors and Arizona AP combined.

Figure 2 also suggests the existence of another threshold at about 80% on the Inventory. This might be regarded as a *threshold for mastery of basic Newtonian concepts.* The figure shows that only students above that 80% threshold are able to score above 80% on the Baseline.

The success of Wells and Swackhamer Honors suggests that this 80% "mastery threshold" is a reasonable goal for high-school physics, even though most university physics falls well short of it. When it is approached, other goals of physics instruction will be much easier to attain.

Acknowledgments

References

1. D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *Phys. Teach.* 30, 141 (1992).
2. R. Thornton and D. Sokoloff, "Learning motion concepts using real-time microcomputer-based laboratory tools," *Am. J. Phys.* 58,858 (1990).
3. F. Reif and S. Allen, "Interpreting and teaching scientific concepts," *Berkeley Cognitive Science Report* No.62, (1990).
4. F. Reif, "Instructional design, cognition, and technology: Applications to the teaching of scientific concepts," *J. Res. Sci. Teach.* 24,309 (1987b).
5. L. McDermott and R. Lawson, "Student understanding of the work-energy and impulse-momentum theorems," *Am. J. Phys.* 55,811 (1987).

Note added in proof
Professor Eric Mazur of Harvard University has just completed the first stage of a pedagogical experiment with some striking results that can be reported here. Mazur was greatly disturbed by the test scores of his students reported above, so he drastically altered his method of instruction in the 1991 fall semester. To focus student attention on important concepts and stimulate active student thinking in his large classes, he changed his methods as follows:

Procedure
Divide the class hour into 15-minute chunks, beginning with a 10-minute lecture that ends with a short qualitative question about the topic under study. Students then have one minute to think about the question, after which they have two minutes to justify their answers with a neighbor. In the last two minutes the instructor guides the class to a resolution of the question. This is the main way that the 1991 course differed from Mazur's previous courses in which he mostly lectured for the entire class hour to passive students. Of course, high-school classes would require a different approach.

Results
  (a) The enthusiastic engagement of the students was obvious in their appreciative response to the classroom interactions.
  (b) *Average post test scores* on the inventory and Baseline tests for the 1991 class (223 students) are 85% and 72%, respectively, a clear improvement over the 1990 scores of 77% and 66% reported above.
  (c) *Exam scores* confirm that this improvement is a real effect. The 1991 class was given the identical final exam taken by Mazur's 1985 class (144 students), achieving a class average of 69.4% compared with 62.9% in 1985. This result is surprising because the 1991 course is required for biology majors, whereas the 1985 course was an elective, so that student population was believed to be better. Furthermore, Mazur regards the problems on that final exam as difficult. Yet the good scores were achieved *without any* problem-solving examples in the lectures. This agrees with our conclusions that students derive little benefit from watching a teacher solve problems.
  (d) The *class Inventory pretest score* was 68%, and the sample is large enough to justify our assumption that this score is probably an equally good descriptor of the 1990 class. Accordingly, we find a strong inventory pretest-post-test gain of 17% for the 1991 class compared with a modest 9% gain for the 1990 class. Most of the gain was at the lower end of the distribution. This is shown by the fact that on the pretest 36% of the students were below the 60% score (which we have identified as a threshold for Newtonian understanding), whereas on the post test only 4% were below threshold. Furthermore, no one failed the course, an unusual result!
  The fairly high pretest mean of 68% is partly attributable to the fact that only 5% of the students were freshmen, but not, evidently, to a background in high-school physics. The II students (no freshmen) who had not taken high-school physics also had a pretest mean of 68%. This is indicative of what bright students can learn without formal instruction.