Published in: *The Physics Teacher*, Vol. **30**, March 1992, 141-158

Force Concept Inventory

David Hestenes, Malcolm Wells, and Gregg Swackhamer

Every student begins physics with a well-established system of commonsense beliefs about how the physical world works derived from years of personal experience. Over the last decade, physics education research has established that these beliefs play a dominant role in introductory physics. Instruction that does not take them into account is almost totally ineffective, at least for the majority of students.

Specifically, it has been established that[1] (1) commonsense beliefs about motion and force are incompatible with Newtonian concepts in most respects, (2) conventional physics instruction produces little change in these beliefs, and (3) this result is independent of the instructor and the mode of instruction. The implications could not be more serious. Since the students have evidently not learned the most basic Newtonian concepts, they must have failed to comprehend most of the material in the course. They have been forced to cope with the subject by rote memorization of isolated fragments and by carrying out meaningless tasks. No wonder so many are repelled! The few who are successful have become so by their own devices, the course and the teacher having supplied only the opportunity and perhaps inspiration.

Table I. Newtonian Concepts in the Inventory.

	Inventory Item
0. Kinematics	
Velocity discriminated from position	20E
Acceleration discriminated from velocity	21D
Constant acceleration entails	
parabolic orbit	23D, 24E
changing speed	25B
Vector addition of velocities	(7E)
I. First Law	
with no force	4B, (6B), 10B
velocity direction constant	26B
speed constant	8A, 27 A
with cancelling forces	18B,28C
2. Second Law	
Impulsive force	(6B), (7E)
Constant force implies constant acceleration	24E, 25B
3. Third Law	
for impulsive forces	2E, llE
for continuous forces	13A, 14A
4. Superposition Principle	
Vector sum	19B
Cancelling forces	(9D), 18B, 28C
5. Kinds or Force	
5S. Solid contact	
passive	(9D), (12 B,D)
Impulsive	15C
Friction opposes motion	29C
5F. Fluid contact	
Air resistance	22D
buoyant (air pressure)	12D
5G. Gravitation	5D, 9D, (12B,D), 17C, I8B, 22D
acceleration independent of weight	1C, 3A
parabolic trajectory	16B, 23D

This gloomy assessment is not intended as a wholesale indictment of the many dedicated and competent physics teachers. It does tell us, though, that effective instruction requires more than dedication and subject knowledge. It requires technical knowledge about how students think and learn. The purpose of this article is to supply some of that technical knowledge and an instrument to help teachers probe and assess the commonsense beliefs of their students. The good news is that this can make a difference! The bad news is, there are no quick fixes!

The central concept of Newtonian mechanics is *force*, so we have designed an instrument to probe student beliefs on this matter and how these beliefs compare with the many dimensions of the Newtonian concept. A copy of the instrument, the *Force Concept Inventory*, is included here for teachers to use in any way they see fit. In the body of the article we discuss the design of the instrument, how to use it, and results obtained with it so far. The instrument has proven valuable at every level of introductory physics instruction from high school to Harvard University. We present extensive baseline data that can be used to assess the effectiveness of physics instruction at any of these levels.

The Inventory data provide a clear, detailed picture of the problem of commonsense misconceptions in introductory physics. It confirms the unanimous conclusion of educational researchers that the problem is very serious. We conclude with a discussion of what can be done about it.

I. Structure and Interpretation of the Inventory

The Force Concept Inventory (see Appendix) requires a forced choice between Newtonian concepts and commonsense alternatives. Table I classifies the Newtonian concepts probed in the Inventory, along with the Inventory items in which they appear. These items are the "correct" Newtonian answers to the Inventory questions. With the exception of question 12 (explained below), there is only one of these answers to each question.

All the concepts in Table I are essential to the Newtonian force concept. The table is best interpreted as a decomposition of the force concept into six conceptual dimensions. All six are required for the complete concept. The kinematics dimension, for example, is essential because the Second Law presupposes the acceleration concept. Physics teachers need no explanation for the rest of the table. Note, though, that each dimension is probed by questions of more than one type.

The first impression of most physics professors is that the Inventory questions are too trivial to be informative. This turns to shock when they discover how poorly their own students perform on it. It is true that the Inventory questions avoid the real complexities of mechanics. But such "trivial questions" are more revealing when they are missed. The Inventory questions are only probes for Newtonian concepts, so one should not give great weight to individual items. There are occasional false positives in the responses of non-Newtonians and false negatives from Newtonians. But only a true Newtonian generates a consistent pattern of Newtonian choices with an occasional lapse at most. Thus, the Inventory as a whole is a very good detector of Newtonian thinking.

As a rule, "errors" on the Inventory are more informative than "correct" choices. The commonsense alternatives to the Newtonian concepts are commonly labeled as *misconceptions.* They should nevertheless be accorded the same respect we give to scientific concepts. The most significant commonsense beliefs have been firmly held by some of the greatest intellectuals in the past,[2] including Galileo and even Newton.[3] Accordingly, these commonsense beliefs should be regarded as reasonable hypotheses grounded in everyday experience. They happen to be false, but that is not always so easy to prove, especially if they are dismissed without a hearing as ill conventional instruction. The Inventory, therefore, is not a test of intelligence; it is a probe of belief systems.

Table II contains a taxonomy of commonsense misconceptions probed by the Inventory. A more detailed taxonomy has been described elsewhere,[2] so we can be brief without attempting completeness. The table lists 28 distinct

Table II. A Taxonomy of Misconceptions Probed by the Inventory. Presence of the misconceptions is suggested by selection of the corresponding Inventory Item.

	Inventory Item
0. Kinematics	
Kl. position-velocity undiscriminated	208,C,D
K2. velocity-acceleration undiscriminated	20A; 21B,C
K3. nonvectorial velocity composition	7C
1. Impetus	
I1. impetus supplied by "hit"	9B,C; 22B,C,E; 29D
I2. loss/recovery of original impetus	4D; 6C,E; 24A; 26A,D,E
I3. impetus dissipation	5A,8,C; 8C; 16C,D; 23E; 27C,E; 29B
I4. gradual/delayed impetus build-up	6D; 8B,D; 24D; 29E
I5. circular impetus	4A,D; 10A
2. Active Force	
AFl. only active agents exert forces	11B; 12B; 13D; 14D; 15A,B; l8D; 22A
AF2. motion implies active force	29A
AF3. no motion implies no force	12E
AF4. velocity proportional to applied force	25A; 28A
AF5. acceleration implies increasing force	17B
AF6. force causes acceleration to terminal velocity	17A; 25D
AF7. active force wears out	25C,E
3. Action/Reaction Pairs	
AR1. greater mass implies greater force	2A,D; 11D; 13B; 14B
AR2. most active agent produces greatest force	13C; 11D; 14C
4. Concatenation of Influences	
CI1.largest force determines motion	l8A,E; 19A
CI2. force compromise determines motion	4C, 10D; 16A; 19C,D; 23C; 24C
CI3. last force to act determines motion	6A; 7B; 24B; 26C
5. Other Influences on Motion	
CF. Centrifugal force	4C,D,E; l0C,D,E
Ob. Obstacles exert no force	2C; 9A,B; 12A; 13E; 14E
Resistance	
R1. mass makes things stop	29A,8; 23A,B?
R2. motion when force overcomes resistance	28B,D
R3. resistance opposes force/impetus	28E
Gravity	
G1. air pressure-assisted gravity	9A; 12C; 17E; 18E
G2. gravity intrinsic to mass	5E; 9E; 17D
G3. heavier objects fall faster	1A; 3B,D
G4. gravity increases as objects fall	5B; 17B
G5. gravity acts after impetus wears down	5B; 16D; 23E

misconceptions along with corresponding Inventory items that suggest their presence when selected. They have been grouped into *six major commonsense categories,* which correspond as closely as possible to the six major Newtonian concepts (or concept dimensions) in Table I. Each commonsense category contains a set of misconceptions about the corresponding Newtonian concept. It will be instructive to discuss each category in turn.

0. Kinematics

In kinematics it is not really appropriate to speak of commonsense misconceptions. Rather, the typical commonsense concept of motion is vague and undifferentiated. Accordingly, as indicated in the Kinematics category in Tables I and II, the Inventory probes for the ability to distinguish between position, velocity, and acceleration, as well as to recognize the vectorial nature of velocity and acceleration. The most rudimentary concept of acceleration is "to know one when you see one."

1. Impetus

Commonsense beliefs tend to be metaphorical and vague with situation-dependent meanings. This is reflected in the use of language. Thus terms like "force," "energy," and "power" are often used interchangeably, as are the terms "velocity" and "acceleration." Even so, most commonsense thinkers distinguish two kinds of force, which we will refer to as *impetus* and *active force*. The term "impetus" dates back to pre-Galilean times before the concept was discredited scientifically. Of course, students never use the word "impetus"; they might use any of a number of terms, but "force" is perhaps the most common. *Impetus* is conceived to be an inanimate "motive power" or "intrinsic force" that *keeps things moving*. This, of course, contradicts Newton's First Law, which is why Impetus in Table II is assigned the same number as the First Law in Table I. Evidence that a student believes in some kind of impetus is therefore evidence that the First Law is not understood.

For an object to move it must be *supplied* with impetus, as expressed by commonsense concept I1 in Table II. As expressed by concepts I2, I3, and I4, impetus can be gained or lost in a variety of ways that vary from student to student. Note the underlying "container metaphor" in the impetus concept: Every object is (like) a container that can store a supply of impetus, like a car stores gas, a kind of "go power" to keep it moving. A few students believe in circular impetus (commonsense concept I5) that tends to move objects in circles; they have been known to justify this by a "training metaphor," which holds that objects tend to do what they have been "trained" to do in the past.[2]

2. Active Force

The commonsense concept of active force is closer than impetus to the Newtonian force concept except, as expressed by concept AF1 in Table II, it is attributed only to certain "active agents" (usually living things), and it acts only by direct contact. Active agents are *causal agents*—they have the power to *cause* motion—to create impetus and transfer it to other objects, as when a boy throws a ball. As indicated by category 2 in Tables I and II, active force is the commonsense concept that corresponds most closely to Newton's Second Law. The commonsense notion closest to a "causal law" is expressed by the syllogism:

Every effect has a cause.
Motion is an effect.
Therefore, motion has a cause.

This leads to the commonsense concept AF2 (motion implies active force).

The vague commonsense analog of the Second Law is that active forces produce motion. When velocity and acceleration are not discriminated as descriptors of motion, it is to be expected that the concept "velocity is proportional to force" (commonsense concept AF4) is not distinguished from "acceleration is proportional to force." Active agents have their limits: a limited capacity to produce motion and a tendency to wear out, as expressed by concepts AF6 and AF7. Note the metaphor of an "acting person" for an active force.

As a technical point, it will be noted that the commonsense belief AF1 (only active agents produce forces) is not evident in the choices A and Bin

question I5. However, we listed it as so in Table II, because to justify those choices in interviews, students appealed to AF1.

3. *Action/Reaction Pairs*

Students often interpret the term "interaction" by a conflict metaphor." They see an interaction as a "struggle between opposing forces." It follows from the metaphor that "victory belongs to the stronger." Hence, students find Newton's Third Law unreasonable, and they prefer some version of the *dominance principle:* In a conflict, the "more forceful" exerts the greater force. Here "more forceful" can mean "bigger," "greater mass," or "more active," as in commonsense concepts ARI (greater mass implies greater force), and AR2 (most active agent produces greatest force).

Because of its strong metaphorical base, the dominance principle (though it is seldom clearly articulated) is so natural to students that it is one of the last misconceptions to be, overcome in the transition to Newtonian thinking. Indeed, it is still to be found in some physics graduate students, as noted in Section III.

4. *Concatenation of Influences*

Common sense offers a number of alternatives**,** as shown I in category 4 of Table II, to the Newtonian force superposition principle. Students often apply the dominance principle to the composition of two forces acting on the same object, with one force winning out over the other. Indeed, they often confuse action/reaction pairs with the superposition of oppositely directed forces on a single object. This is another example of poorly differentiated concepts so typical of commonsense thinking.

5. *Other Influences on Motion*

Unlike the Newtonian world, the world of common sense does not have a unitary concept of force. Besides active forces, there are other influences on motion, as listed in category 5 of Table II. Actually, the Inventory does not contain any items designed specifically to probe for the centrifugal force misconception listed in the table. That misconception is only suggested by the form that the listed items take in the questions. Verification would require an interview or explanation from the students. We have encountered high-school physics teachers who think that centrifugal force is a distinct kind of force. Such is the power of a name!

In the world of common sense, *obstacles* like chairs and walls do not exert forces, "they just get in the way." Mass is regarded as a kind of resistance, because it "resists" the efforts of an active agent. Motion occurs only when the active force "overcomes" the resistance (note the metaphor), and it ceases when the force becomes "too weak."

In the world of common sense, "gravity" is not necessarily the same as "gravitational force." When they are the same, the commonsense concept G3 (heavier objects fall faster) can be regarded as a special case of AF5 (acceleration implies increasing force). Concept G3 may appear to be true, but the underlying misconception is a matter of scale, to which common sense is often oblivious. It is believed that gravity varies significantly over a few meters, whereas the variation is actually about one part in 10^{13}.

The belief G1 that air pressure contributes to gravity is common only among very naive students. Among other things, question 12 was designed to detect this misconception. The fact that the net force due to air pressure is actually upward (buoyant force) instead of downward, was hardly recognized by students at any level, for item 12D was very rarely selected. Interviews of 16 graduate students revealed that only two of them really understood the buoyant force concept. A third of the others could state Archimedes principle, but they did not know that the buoyant force is due to a pressure gradient, and some offered very peculiar hypotheses to explain it. No doubt this sorry state of affairs is largely due to the fact that buoyancy gets little attention in the physics curriculum today. Because of all this, item 12D is not very informative, and we allow 12B as an acceptable Newtonian choice. We have retained item 12D, nevertheless, because it is such a good pretext to interview students about buoyant force. Besides, some teachers might think physics students should know why things float!

II. Results and Implications

The *Force Concept Inventory* test has been given to more than 1500 high-school students and more than 500 university students. Results are displayed in Table III along with post test scores on the *Mechanics Baseline,* described in a companion paper.[4] For the purpose of rough comparison, the Baseline test can be regarded as a problem-solving test involving basic Newtonian concepts. Except for two of the authors (Wells and Swackhamer), all teachers with class test results in Table III were blind to both tests when their teaching was done. Both Wells and Swackhamer were scrupulously careful not to teach to the tests in their own classes.

Table III. Inventory and Baseline Scores.

Class	Inventory Pretest % (S. Dev.)	Inventory Post test % (S. Dev.)	Baseline Post test % (S. Dev.)	Post test Number of Students N
High School				
Arizona Reg.	27 (11)	48(16)	32 (11)	612
Wells Reg.	28 (14)	64 (20)	42 (16)	18
Chicago Reg.	27	42		56
Arizona Hon.	33 (13)	56 (19)	37 (15)	118
Wells Hon.	42 (18)	78 (15)	62 (17)	30
Swackhamer Hon.	28	66	47	63
Arizona AP	41 (16)	57 (18)	39 (15)	33
Swackhamer AP	73	85		11
University				
Van Heuvelen 105	34 (14)	63 (18)	61 (18)	116
Wells 105	36	68	43	44
Arizona State Reg.	52 (19)	63 (18)	48 (15)	139
Harvard Reg.		77 (15)	66 (14)	186
Harvard Honors			73 (11)	75

Remarks: Mean scores (%) and standard deviations on all tests are given in percent. N is the number of students taking the post test; variations in the numbers taking pre- and post tests were judged to be insignificant or, at least, uninformative. Arizona Reg. combines data from 15 teachers. Chicago Reg. are regular high-school classes in the Chicago area. The Chicago Reg. teacher employed neither the Wells nor the A. Van Heuvelen teaching methods.

High-School Results

Besides Wells and Swackhamer, eighteen Arizona high-school physics teachers participated in the study. Mean scores for all their students combined are given in Table III, grouped according to level: Arizona Reg. and Arizona Hon. denote first-year Regular and Honors physics, respectively. Arizona AP denotes a second-year "Advanced Placement" physics course, which usually uses a university calculus-based physics textbook. On an elementary math (mostly algebra) test, the three levels are distinguished by percent mean scores (standard deviations) of 40 (19), 53 (22), and 63 (20), respectively. However, in agreement with previous conclusions,[1] this has no significant correlation with the data in Table III, except possibly the slightly higher Arizona AP score on the Baseline. Since the initial math scores for the Wells Reg. and Wells Ron. are about the same as the average for the Arizona schools, but the post test physics scores are much higher, we conclude that math background is not a major factor in the high-school results in Table III.

Table IV Arizona Inventory Scores vs Teacher Competence Ranking

Teacher Competency Ranking	Regular Pre %	Regular Post %	Honors Pre %	Honors Post %	AP Pre %	AP Post %	Socio-Econ. Level of School	Number of Students
1	28	48					4	50
2	24	45	30	52	56	64	5	7/26/8
3	29	53					3	35
4	25	44					1	93
5	34	59					3	18
6			39	67			2	46
7	25	52					3	67
8	29	46					3	73
9	28	40	25	40	39	47	4	32/36/9
10					35	60	2	16
11	30	51					2	42
12	28	64					5	15
13	25	49					3	12
14			37	73			2	10
15	27	50					2	75
16	31	47					1	45
17	24	44					1	12
18	23	33					5	26

Remarks: Teacher competency ranking #1 corresponds to the most competent teacher. Socioeconomic level #1 corresponds to the school with the highest socioeconomic student populations.

The overall gains of 20% for Arizona Reg. and 23% for Arizona Ron. are certainly significant, but we had reason to hope for more. All the teachers are involved in an NSF physics education project conducted by Wells and Hestenes. In the first year of the project, pretest/post test data were gathered for the classes

of each teacher. During the following summer, the teachers attended an intensive six-week workshop where they were introduced to a new method for teaching high-school physics developed by Wells, which they all agreed to try out in the following year. When pretest/post test data for the second year were compared with the first-year data for each teacher, a significant improvement was evident in only a couple of cases. Overall post test data for the two years differed by only a few percent, scarcely greater than variations in the pretest data. We reluctantly concluded, therefore, that no overall improvement was achieved, so we recorded the combined data for both years in Table III to form a massive reference database. Table III also includes data from a single, conventionally taught "regular physics" course at Swackhamer's school in Chicago to show that the results are not better than the Arizona results. We have no reason to suspect that better results will be found at typical high schools anywhere in the United States.

Pretest/post test data for classes of the individual teachers are displayed in Table IV, along with a rough socioeconomic ranking of their schools on a five point scale: (1) wealthy, (2) upper middle, (3) middle, (4) lower middle, and (5) low. Some of the schools at the lowest levels have substantial numbers of native Americans and Hispanics, but few of these students take physics. The teachers are designated and ordered by a competence ranking with 1 as the highest. The competence ranking is fashioned from a subjective combination of academic background, mechanics diagnostic score, and teaching experience. Eight of the teachers have a B.A. in physics or a master's in science education with considerable physics. Fourteen have scores on the Mechanics Diagnostic above 80%, roughly comparable to the same score on the Inventory (see Section III). In our companion paper,[4] we give reasons for regarding 80% on the Inventory as a threshold score for Newtonian thinkers. Accordingly, we conclude that at least 80% of the Arizona teachers are well qualified to teach high-school physics. Actually, we are impressed with the performance of one of the lower ranked teachers with minimal physics background, who (and we think this is important) has had the benefit of working closely with a highly ranked teacher at the same school.

Perusing Table IV, we see no correlation of scores with socioeconomic level, and computation of average scores for each level confirms this. One reason for this result is that the subset of students who take physics is usually not typical of the student population at the school. However, there is great variation from school to school. For example, the students of teacher 12 were bright and motivated, some of them children of engineers and teachers in a mining community. Of teacher 1's students, 60% were female and nearly 20% were native American. In the school of teacher 9, the student population is about equally divided between black, Hispanic, and white, but 90% of those in physics were white and discipline was difficult. In this light, the independence of post test scores with socioeconomic level is all the more remarkable. The data in Table IV also show no correlation of post test score with competence level, with the exception of teacher 18 (who had a Diagnostic score of only 39%). This suggests that student scores are unlikely to surpass the teacher's score! However, that conclusion is confounded by the fact that teacher 18's school is at the very bottom of the socioeconomic ladder, with many children of migrant farm workers, who have very low scores on every kind of academic test.

The Arizona scores in Table III should be compared with the much better scores in the Wells Honors and Swackhamer AP classes. The contrast is all the greater considering that we have reason to doubt the validity of the highest Arizona Honors post test scores (67 and 73) in Table IV. Note that the Arizona AP scores in Table III do not exceed the Arizona Honors scores despite the extra year of physics, and the average AP Inventory gain is a mere 5%. In contrast, for the Wells Honors the Inventory gain is an impressive 36%, and the combined post test scores are comparable to those for Harvard University Reg., a Harvard calculus-based introductory physics course for science majors (mostly biology and premed). Wells has consistently achieved similar results for several years.

The greater achievement of Wells Honors compared with Wells Reg. is noteworthy. The perception of both Wells and Swackhamer is that the two classes do not differ greatly in intelligence and mathematical competence. The main difference is in *attitude.* Students in Honors physics are highly motivated and eager to pursue class activities on their own. In contrast, students in Regular physics require continual teacher supervision. The socioeconomic level for both Wells and Swackhamer classes is upper-middle class. Most families of professionals are included in that level, and that is probably the largest source of motivated students for Honors physics.

The high Swackhamer scores in Table III are at least partly explained by the fact that Swackhamer had much more intensive training in the "Wells method" than the Arizona teachers. Besides having prior familiarity with key ideas of the method, Swackhamer spent the better part of an academic year working closely with Wells in his classroom. The Swackhamer AP scores should be compared with the Harvard University Reg. scores, since 94% of the latter had also taken a year of high-school physics.

The Wells and Swackhamer data establish conclusively that very large gains in overcoming misconceptions and understanding Newtonian mechanics are possible in high-school physics. Comparison with the Arizona data strongly supports the conclusion that such gains are not possible with conventional instruction. The question remains, why was there so little improvement in the Arizona results after the workshop on the Wells method?

An answer to that question is suggested by examining what happened in the workshop. All the Arizona teachers were excited about the workshop and the Wells method, and they will testify that it has greatly improved student interest and the quality of their instruction. The method is computer-based and laboratory-oriented instruction with no lectures, but with much class discussion and some special techniques to stimulate it. The computers have the advantage of reducing the busy work in data collection and analysis, so more time can be devoted to understanding what it all means. From discussions with the teachers after the second year, it has become clear that they were so involved with the mechanics of the method-computers, lab activities, discussion technique-that they failed to fully appreciate the crucial pedagogical core that makes it effective. The net result is another demonstration that *technology by itself cannot improve instruction.* The best that technology can do is enhance the effectiveness of good pedagogy.

University Results

As part of a pedagogical experiment, Professor Alan Van Heuvelen of New Mexico State University visited Hestenes at Arizona State University for two consecutive fall terms to test the effectiveness of new pedagogical techniques in

Arizona's Physics 105. This is not the place to discuss those techniques; suffice it to say that they have much in common with those employed by Wells. Physics 105 is an "interface course" intended to prepare students who have not taken high-school physics, or are otherwise academically deficient, to take the calculus-based university physics (called Arizona State Reg. here).

The effectiveness of Van Heuvelen's approach becomes evident in Table III by comparing his course, Van Heuvelen 105, with Arizona State Reg., and Harvard University Reg., both taught by conventional methods. Note that the Inventory pretest score is not much better than that of the high schools. By this and other indices, such as math background, the 105 students as a group would surely fall into the lowest third of students taking Arizona State Reg. Yet the post test scores are clearly superior. Here again we have clear evidence that pedagogy can make a difference!

To expand the pedagogical experiment with Physics 105, Wells tried teaching the course for one semester. It must be said that the Wells method is not compatible with the large lecture class format required for Physics 105, and this is reflected in the comparatively low Baseline scores shown in Table III. However, improvements in the laboratory portion of the course were considerable, and this is reflected in the Inventory post test score.

The Harvard data in Table III provide a valuable index of the best that can be expected from conventional instruction in university physics. Students in the Harvard University Honors are mostly physics majors, but the course is essentially the same as Harvard Reg.

The time available to take the tests is an important variable. For the Harvard students the tests were administered on computers, and they could spend as much time as needed on each problem. However, the computer would not allow back-tracking. Some students reported realizing that they had made a mistake on a previous problem, which they could not backtrack to correct. For this reason the Harvard scores may be slightly low.

The Harvard computers accurately measured the time spent by each student on each problem. From this we have an average Harvard time of 23 minutes for the Inventory and 40 minutes for the Baseline. All the high-school classes in Table III had ample time for both tests, up to a full class period (50 minutes). The three Arizona State classes were allowed 40 minutes for the Inventory, and the Baseline was included in the final exam, for which ample time was available.

III. Test Validity and Item Analysis

The *Force Concept Inventory* was designed to improve on the *Mechanics Diagnostic* test described in detail elsewhere.[1] The results originally obtained with the Diagnostic have since been replicated many times by others, so we have great confidence in the reliability of the test and the conclusions drawn from the data. Further confirmation comes from the Inventory results in Table III. Indeed, the percentage scores on both tests seem to be quite comparable measures of Newtonian conceptual understanding. The pretest/post test Inventory scores of 52/63 for Arizona State Reg. are nearly identical to the 51/64 scores obtained with the Diagnostic for the same course. Moreover, besides the data in Table III, we have post test averages of 60 and 63 for two other professors teaching the same course. Thus, we have the incredible result of nearly identical post test scores for *seven different professors* (with more than a thousand students). It is hard to

imagine stronger statistical evidence for the original conclusion that Diagnostic posttest scores for conventional instruction are independent of the instructor. One might infer from this that the modest 11% gain for Arizona State Reg. in Table III is achieved by the students on their own.

Though percentage scores on the Inventory and the Diagnostic are of comparable significance, the Inventory has the advantage of supplying a more systematic and complete profile of the various misconceptions, as delineated in Table II. About half the questions in the Inventory are essentially the same questions in the Diagnostic, because we could not find better ones to replace them. Considerable care was taken to establish the validity and reliability of the Diagnostic.[1] Formal procedures to do the same for the Inventory are unnecessary because the test designs are so similar and such diverse data are presented here. Nevertheless, we took the precaution of interviewing students about their responses to the Inventory questions. One of us, Swackhamer, interviewed 20 students in Wells's classes. He was amazed at how predictable the responses were, as if the students were reciting the results of previous interviews.[2] He found that students had firm reasons for most of their choices, though he detected vacillation among some alternatives. Non-Newtonian choices were rarely made by students with the relevant Newtonian concept, but Newtonian choices for non-Newtonian reasons were fairly common. Therefore, except possibly for high scores (say, above 80), the Inventory score should be regarded as an upper bound on a student's Newtonian understanding. All this is in complete accord with conclusions about the Diagnostic.[1,2]

One of us (Hestenes) interviewed 16 first-year graduate students beginning graduate mechanics at Arizona State University. The interviews were in depth on the questions they had missed on the Inventory (more than half an hour for most students). Half the students were American and half were foreign nationals (mostly Chinese). Only two of the students (both Chinese) exhibited a perfect understanding of all physical concepts on the Inventory, though one of them missed several questions because of a severe English deficiency. These two also turned out to be far and away the best students in the mechanics class, with near perfect scores on every test and problem assignment. Every one of the other students exhibited a deficient understanding of buoyancy, as mentioned earlier. The most severe misconceptions were found in three Americans who clearly did not understand Newton's Third Law (detected by missing question 13) and exhibited reading deficiencies to boot. Two of these still retained the Impetus concept, while the other had misconceptions about friction. Not surprisingly, the student with the most severe misconceptions failed graduate mechanics miserably, while the other two managed to squeak through the first year of graduate school on probation.

Interviews with the graduate students who had difficulty with Newton's Third Law proceeded by asking them to draw free-body diagrams for each vehicle in question 13, as well as for the two-vehicle system as a whole. This revealed a host of deficiencies. All three students were unable to draw correct diagrams; they had difficulties isolating the system of interest, separating external agents from the object, and determining what forces act where. They failed to realize the universality of Newton's Third Law or recognize the circumstances where it applies. Like beginning students, they confused the balance of forces on a single object (superposition) with the equal and opposite forces on different objects in an

interacting pair. They clearly applied the "dominance principle" to the action/reaction pair in question 13. The interviews brought each student to recognize, finally, that failure of the interaction between the vehicles to obey Newton's Third Law would result in a self-accelerating system. The heartening result of the interviews was that all three students could be led to recognize, articulate, and correct each mistake when attention was directed to it Socratically. At last, perhaps, they arrived at a secure understanding of the Third Law.

One disturbing observation from the interviews was that five of the eight Americans, as well as five of the others, exhibited moderate to severe difficulty understanding English text. In most cases the difficulty could be traced to overlooking the critical role of "little words" such as prepositions in determining meaning. As a consequence, we discarded two interesting problems from our original version of the Inventory because they were misread more often than not.

Table V gives the classification for every item in the Inventory keyed to their interpretations in Table I. It also gives the item pretest/post test percentages for the groups of greatest interest in Table III. This data contains a wealth of information, though you have to know something about the teaching to extract much of it. Here are some conclusions from the data. Of particular interest are clues to how the Arizona teachers might become more effective.

(1) Wells and Van Heuvelen successfully addressed Newton's Third Law. The Arizona teachers did not (questions 2, 11, 13, and 14). The key Third Law question is 13; note how different from 14 it appears to most students. Also note the "progress" of Arizona students on question 11, from mistake B to mistake D.

(2) Questions lending themselves to analysis by force diagrams are: 5,9,12,22 for identification of forces, and 18, 28 for finding *net* force. In general Wells and Van Heuvelen did much better on these than the Arizona teachers.

(3)Why do Wells and Van Heuvelen do so much better than Arizona teachers on the "trivial" kinematics question 21? Perhaps because they are more systematic and thorough in teaching graphical and "motion map" techniques for representing motion.

(4) Questions 19 and 29 are weak discriminators, so they could be dropped from the test. Question 19 was intended to test for understanding of the superposition principle, but the high percentage selecting the correct response shows that it failed. On reflection, it is clear that the Newtonian response could as easily be justified by the non-Newtonian dominance principle. Item 9D is more discriminating, though it involves other concepts as well. Similarly, selection of the correct response 29C might be based on an erroneous belief in impetus decay.

(5) The high percentage of students choosing 23A is curious. The choice might be grounded in the perceptual experience of dropping an object out the window of a moving car.

(6) Note Harvard's most popular wrong choice, 24C. That could be the result of sloppy reading—jumping to the conclusion of an impulsive force acting as in 6B. Some students evidently realized the mistake when reading 25, but could not go back to correct it.

(7) Note the persistence of 3B and 3D (heavier falls faster). This should be compared with the much better performance on question 1, which is no doubt closer to what is discussed in most classes.

(8) Finally, it is worth noting that "teaching to the test" or a breach of test security may be revealed in anomalous frequency distributions on the test items, most noticeably when all the students select the same wrong answers.

IV. Uses for the Inventory

The *Force Concept Inventory* is not "just another physics test." It assesses a student's overall grasp of the Newtonian concept *of force.* Without this concept the rest of mechanics is useless, if not meaningless. It should therefore be disturbing rather than comforting that students with only moderate scores on the Inventory may score well on conventional tests and get good grades in physics. Of course, experienced teachers have learned to avoid problems that are "too hard" for the students. That includes most *qualitative problems* that seem so simple until student answers are examined. Students do better on *quantitative problems* where the answer is a number obtained by substitution into an appropriate equation, and even on harder problems that require some algebraic manipulation. So should we not be satisfied that they have developed quantitative skills? After all, physics is a quantitative science! Or do we have here a selection process that directs teachers to problems that students can answer with a minimum of understanding?

Like its predecessor, the *Mechanics Diagnostic,*[1] the *Force Concept Inventory* can be used for both instructional and research purposes. The applications fall in three main categories:

(1) *As a diagnostic tool,* the Inventory can be used to identify and classify misconceptions. It is especially valuable for teachers, to raise their awareness of misconceptions among their own students. The greatest insight is attained from interviews based on the Inventory, where students are asked to give reasons for their choices. Interview techniques for uncovering misconceptions in mechanics have been discussed by McDermott.[5] Interviews are very time consuming, but they need not be repeated with every class, because the misconceptions are universal. Once the teacher gains sufficient insight into misconceptions, interviews are unnecessary for it is known beforehand that the misconceptions are present and must be addressed. The interview technique for individual students should be transformed into a class discussion technique for probing misconceptions and stimulating interaction among the students to induce conceptual change. When skillfully done, this is one of the most effective means of dealing with misconceptions. Arnold Arons is perhaps its most experienced practitioner, and he has sage advice to offer.[6]

(2) *For evaluating instruction,* we now have abundant evidence that the Inventory is a very accurate and reliable instrument. We have collected both pretest and post test data for research purposes, but the pretest scores are so uniformly low for beginning physics students that further pretests are really unnecessary, except to convince diehard doubters or to check out the conceptual level of anew population. The evidence that large Inventory gains are possible is now sufficient for us to conclude that, for effective instruction, *only the posttest score counts.* Pretest/post test gains will be large if the pretest scores were low but small if pretest scores were high. The final result will be nearly the same in either case—if the instruction is effective. It is no longer acceptable to blame low post test scores on poor background of the students. The main deficiency is likely to be in the instruction.

It is possible, of course, to get high scores on the Inventory by "teaching to the test." Students can memorize Inventory answer as well as anything else. But the answer have little significance in themselves. It is the student reasoning to get those answer that really matters. The Inventory questions are probes to stimulate that reasoning, so the process is short circuited if the "correct answers" are supplied to be repeated back without thinking. Inventory questions can be useful starting points of foci for class discussions, but that precludes using the Inventory for assessment, so it is advisable to choose other means for stimulating discussion. However, this is for the teacher to decide, of course!

(3) *As a placement exam*, the Inventory has limited value in high school. It is not a test of ability, so it should certainly *not* be used to place beginning students in, say, Regular or Honors physics. It can be used in colleges and universities to help determine if student understanding of introductory physics is sufficient for a more advanced course. For that purpose it should probably be used in conjunction with the *Mechanics Baseline* or some other test.

V. Overcoming Misconceptions

Knowledge about the nature and extent of student misconceptions is insufficient by itself to improve the effectiveness of instruction. Simply telling students about their misconceptions, like teaching to the test, has very little effect. To induce significant conceptual change, a well-designed and tested instructional method is essential. Like any other complex intellectual skill, effective teaching requires sound technical knowledge.

This is not the place to discuss specific instructional techniques, but some general remarks on our own instructional orientation may be helpful. First, we would like to warn against a piecemeal approach directed at each misconception separately. Misconceptions can be successfully overcome only when something better (namely, Newtonian concepts) is available to replace them. Moreover, one great strength of Newtonian mechanics is that it is a *coherent* conceptual system, and this can have as much impact on student learning as it did on scientists adopting the system in the first place. Accordingly, we aim first at teaching a *unitary concept of force* with all six of its major components listed in Table I. Within this context, student misconceptions are elicited and treated when they are prone to conflict with the Newtonian concepts. The instructor must *anticipate* when the discussion of specific misconceptions is likely to be most profitable, *focus* student attention on the crucial issues, and bring the discussion to a satisfying *closure.* This requires planning, preparation, and practice. It is not easy to do well, but it can be very rewarding for teacher and students alike.

Although instruction must deal with misconceptions systematically to be efficient, in our experience it is unnecessary to deal with every single one of them explicitly. Some minor misconceptions, such as "circular impetus," tend to disappear spontaneously with the treatment of major misconceptions and the growth of Newtonian concepts. Ollef among the major misconceptions we place the *impetus concept of motion* and the Dominance principle or the *conflict concept of interaction.* These are the most difficult and usually the last of the misconceptions to be overcome. Unless dealt with effectively, they may persist in the minds of students for a long time, even into graduate school, as we have already noted. For students who major in physics, all the misconceptions tend to disappear spontaneously through processes of acculturation. This is evident from

the paradoxical fact that few physicists can recall having ever believed, let alone having overcome, any of the misconceptions, though research has established unequivocally that everyone has them before learning physics. Conventional instruction does work for some students, but at best it is slow and inefficient. We now have strong evidence that misconceptions must be taken, into account to improve the efficiency of physics instruction.

But that is not enough by itself. In traditional instruction, problem-solving skill is regarded as the *sine qua non* of physics understanding. We do not quarrel with that, but we wish to emphasize that certain concepts and modes of reasoning must be developed *before* problem-solving instruction can be effective. This includes skills in graphical and diagrammatic representations of motion and forces, critically discussed in Ref. 7 and emphasized in the successful (according to the results of Table III) courses of Wells, Swackhamer, and Van Heuvelen. Our data suggest that there exists a kind of conceptual threshold near 60% on the Inventory. Below this threshold, a student's grasp of Newtonian concepts is insufficient for effective problem solving. This would explain the uniformly low scores of the Arizona Reg. courses on the "problem solving" Baseline test (Table III), for none of them approach 60% on the Inventory.

For beginning students below the 60% threshold, it is especially important to take misconceptions into account. Arons[6] presents the most extensive discussion of this matter, as well as many other insights into physics instruction. Minstrell and Stimpson[8] provide a systematic approach to some of the most basic misconceptions, especially about gravity. Clement[9] has developed an instructional technique called "bridging," which exploits strengths in student intuitions by inducing them to establish conceptual "bridges" between different physical situations, thus sharpening their recognition of similarities and differences.

Acknowledgments

This work was supported by the National Science Foundation. David Marcus was very helpful in collecting and analyzing the data. We thank the many high-school and university teachers who so graciously cooperated in collecting data from their classes, especially Eric Mazur, who supplied the Harvard data and suggested improvements in the Inventory.

References

1. I. Halloun and D. Hestenes, "The initial knowledge state of college physics students," *Am. J. Phys.* **53**, 1043 (1985).
2. I. Halloun and D. Hestenes, "Common-sense concepts about motion," *Am. J. Phys.* **53**, 1056 (1985). This article contains a fairly complete taxonomy of misconceptions about mechanics.
3. M. Steinberg, D.E. Brown, and J. Clement, "Genius is not immune to persistent misconceptions: Conceptual difficulties impeding Isaac Newton," *Int. J. Sci. Educ.* **12**, 265 (1990).
4. D. Hestenes and M. Wells, "A mechanics baseline test," *Phys. Teach.* **30**, 159 (1992).
5. L. McDennott, "Research on conceptual understanding in mechanics," *Phys. Today* **37** (7), 24 (1984).
6. A. Arons, A *Guide to Introductory Physics Teaching* (Wiley, New York, 1990), p. 50ff.
7. D. Hestenes, "A modeling theory of physics instruction," *Am. J Phys.* **55**,440 (1985).
8. J. Minstrell and V. Stimpson," A teaching system for diagnosing student conceptions and prescribing relevant instruction," *Sci. Teach.* (submitted). (Copies are available from J. Minstrell, Mercer Island High School, 9100 SE 42nd, Mercer Island, WA 98040.)
9. J. Clement, "Nonformal reasoning in experts and science students; the use of analogies, extreme cases and physical intuition," in J. Voss, D. Perkins, and J. Segal (eds.) *Informal Reasoning in Science Instruction* (Lawrence Erlbaum Associates, Hillsdale NJ, 1991), pp. 345-362. (Copies are available from J. Clement, Scientific Reasoning Research Institute/Hasbrouck, University of Massachusetts, Arnherst, MA 01003.)