INVESTIGATING ANIMATIONS FOR ASSESSMENT WITH AN ANIMATED VERSION OF THE FORCE CONCEPT INVENTORY

By

MELISSA HAYES DANCY

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

PHYSICS

Raleigh, NC

2000

APPROVED BY:

Dr. John Risley	Dr. John Park
Dr. Stephen Reynolds	Dr. Robert Beichner, Chair of Advisory Committee

BIOGRAPHY

I was born in Charlotte, NC in 1970 and lived there until I graduated from high school in 1988 and went to college. I have always had a strong desire to understand why the world works as it does. This desire, and a love of learning, undoubtedly led me into a physics major at Furman University. My education at Furman was broad, which I am thankful for. There I learned how to learn and to believe in myself. I graduated with my B.S. in physics in 1992 and entered Purdue University's department of physics as a graduate student. During my first year I was given the opportunity to teach recitation sections. I thought teaching would be easy, but it turned out to be one of the greatest challenges I had ever faced. Although I felt I was doing an excellent job, and my students agreed, I knew that my students were not learning as much as they should. It was during that year that I began my life's quest to learn more about how people learn and to improve physics instruction. Over time, I became less interested in traditional areas of physics and increasingly interested in physics education.

I received my Masters degree in physics in 1994 and began work on a degree in science education at Purdue. In 1996 I completed my second Masters degree and came to NC State because I felt NC State offered me more for the long term. I was absolutely correct. The courses I took here were excellent, and the support of a top research group and advisor have been invaluable.

In 1997 I married Tom Lipinski who has been extremely supportive of my work. Our son, Logan, was born in the spring of 1998. He has taught me the power of focus, and balance in life.

After completing my Ph.D. I will join the physics faculty at Davidson College for two years to continue work related to this thesis.

ACKNOWLEDGEMENTS

This project was completed because of the inspiration, encouragement, advice, and physical support I received from so many people who were willing to share their time and energy. The idea for the project was directly inspired by the work of Aaron Titus. I am also grateful to Aaron for providing me with emotional encouragement, and an intellectual sounding board as the project progressed. The Physlets, which drive the animations, were possible because Wolfgang Christian is willing to so freely share his work. The authors of the FCI provided the questions for analysis and I am thankful to them for allowing me to distribute and publish part of their work. The cgi script that was used to collect student answers was essentially written and implemented by Larry Martin and Brian Pittman.

I owe many thanks to Robert Beichner, Mario Belloni, Abebe Kebede ,Tom Lipinski, John Merrill, Todd Leif, The Ohio State Physics Education Group, George Parker, and Aaron Titus for their help with the data collection. I was very fortunate that so many people were willing to help me access students.

During my entire time at NC State I was supported by a NSF Fellowship for instructional technology. The Fellowship allowed me to focus on my own progress, enabled me to attend numerous national meetings, and gave me access to funds which made the data collection and analysis process much easier. I am very grateful to have been able to receive such support.

I am also grateful to my research group, David Abbott, Rhett Allain, Scott Bonham, Jeff "the walking database" Saul, and Duane Deardorff for providing a sounding board for ideas and a strong day to day support system. My committee members gave me lots of helpful feedback. And I am particularly grateful to have had an advisor, Robert Beichner, who was willing to let me follow my own ideas and was always so positive and encouraging.

Finally, I am fortunate to have a very loving and supportive husband, Tom Lipinski, whose patience and encouragement have been a pillar of strength for me. He has so kindly listened to me when I needed to vent, motivated me when I needed a kick, and shared in my accomplishments with joy and excitement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

DANCY, MELISSA HAYES. **Investigating Animations for Assessment with an Animated Version of the Force Concept Inventory**. (Under the direction of Robert J. Beichner)

Due to new technologies, it is now possible to assess students' conceptual understanding using computer animation. However, very little research has been done to guide the use of animations for assessment. I have looked at this area by replacing static pictures and descriptions of motion with animations on the Force Concept Inventory. The animated and traditional versions were then given to hundreds of students. Data on ACT scores and gender were also collected for many of these students. It was found that the animation can affect the answers students give, especially when the question involves motion and the animation is central to answering the question. I also found that performance on the animated questions is less correlated with verbal ability than performance on the traditional questions.

The second phase of this project involved think-aloud interviews with students to determine how the animations affected their interaction with each question. The results indicate that the animation can reduce misunderstandings due to reading difficulty or question vagueness. The animation can also help students to visualize situations.

For these reasons, the animated questions were often found to be superior to the traditional question because of increased validity. Based on the results of this study, I believe that computer animation can provide a valuable, and often improved method of assessment.

Link to Online Physlet Force Concept Inventory

Return to Homepage

Chapter 1
Introduction

1.1 Purpose

Assessment is an integral component of the learning process. Traditionally, large-scale assessment of student understanding has been accomplished through the use of pencil and paper. However, it is now possible to assess students in alternative ways through the use of technology. Unfortunately, there is no strong research base to guide teachers in the use of technology for assessment.

In theory, the computer should be an asset to learning and assessment. After all, it can engage students interactively one on one, it can present movement, graphics and sound, it can easily record student input, and it can facilitate communication that would otherwise be difficult or impossible. Unfortunately, the computer by itself does not educate. As Rieber[1] put it, asking if the computer is a good tool is a bit like asking, "is a hammer is a good tool to use? The answer is sure, sometimes, but it all depends. It's great for hammering and pounding nails, but pretty lousy for cutting hair." There are times when the computer can be used to facilitate learning and times when it is just an expensive waste of time, or even worse, damaging to learning.

Consider the report of Wenglinsky[2] who analyzed 4th and 8th grade mathematics scores on the National Assessment of Educational Progress along with teachers' reports about classroom technology use. He found that students' mathematical achievement went down as computer use in school increased. This result seems to indicate that the computer does not benefit learning.

However, Wenglinsky also considered the way students were using the computer in school. He found that the low scores could be accounted for by the widespread use of drill and practice computer use. Students whose teachers primarily used the computer for "drill and practice" performed nearly half a grade level worse than students who did not use the computer in this way.

Encouragingly, he did find evidence that computer time could benefit students. Using the computer for "simulations and applications" was associated with learning gains. Unfortunately, only 7.5% of fourth-grade teachers and 27.2% of eighth-grade teachers reported this type of computer use as their primary application.

This results do not definitively show that drill and practice will lead to decreased performance and simulations and applications will increase learning since it is correlational data only. For example, teachers may be more inclined to give lower ability students drill and practice. However, Wenglinsky's results are intriguing.

The computer can be used effectively or misused in the classroom. In and of itself, it is not a solution of all educational problems, nor is it the undermining of education as some have feared. But how are teachers and instructional designers to know which uses will be valuable? Although there is a research base (as will be discussed in the next chapter), it is incomplete. The purpose of this project is to add to this body of research, particularly in the area of assessment.

1.2 Research Questions

In this project, I have chosen to focus on a very specific use of the computer, the use of computer animations in conceptual assessment. The specific questions that will be addressed follow.

Question 1

When static pictures and descriptions of motion are replaced with computer animations, how are students' responses to Force Concept Inventory[3] *questions affected?*

It seems logical that computer animations, instead of static pictures and verbal descriptions, will affect student answers to conceptual questions. When students interact with an animation it requires them to recognize motion in real time. In order to answer a traditional question they must use their verbal skills and perhaps, interpret meaning about motion from a static picture. The cognitive skills involved in the two types of questions are different, so an animation effect would be logical. At first glance, it also seems reasonable that animation would only improve student performance since the ability to recognize motion seems to be lower on the cognitive ladder than the ability to interpret a representation of motion.

Answering this question entails not only looking to see *if* students answered differently on the animated version of the question, but also *how* their answers changed. In other words, was the animation always helpful? Also, it is important to look for categories of responses from which generalizations can be made. For example, is there an effect based on such things as physics concept, interaction with animation, gender, or general intellectual ability? It is possible that certain groups may be affected by the animation more than others. For example, students with less verbal ability may benefit from a less verbal test.

Question One is the subject of Chapter Three of this dissertation. It is answered through a statistical analysis of a large data set.

Question 2

If student responses are affected by the animation, what are the reasons behind these effects?

This is both the most important and difficult question to answer. If this research is to be of value beyond the scope of the FCI then the reasons behind the effects must be investigated. However, questions of "what" are much easier to answer than questions of "why". Question two is primarily answered through a qualitative analysis of think-aloud interviews discussed in Chapter Four.

Question 3

What implications can be drawn about the value of traditional and animated versions of the Force Concept Inventory for conceptual assessment of physics concepts?

The FCI is given to thousands of students every year and used as an assessment of instruction. Can animations be used to increase the validity or reliability of the test? If the effect of animation is different for different groups of students, what are the implications for assessment? Also, those in the field of physics education are not in agreement about the value of the test or what it actually measures. (See chapter 2) Perhaps the animated FCI can add to our understanding of these issues.

Question Three is answered through a combination of quantitative and qualitative methods. A test item analysis (See Chapter Three) was performed for both versions of the FCI and interview data (See Chapter Four) was used to add additional insights.

1.3 Rationale

Why investigate computer animations in conceptual assessment? Well, as was discussed at the beginning of this section, there is a need for understanding more about educational uses of computer technology because of the lack of such research, and because of the increasing presence of the computer in the classroom. Billions of dollars are being directed into educational uses of technology. We need to know how to make this expenditure worthwhile. Although some research has been done on educational applications of computer technology, almost none has looked at using animations for assessment. The computer has been widely used as a homework delivery system (e.g.

WebAssign[4]) of traditional problems or to alleviate grading demands on the instructor, but little has been done to integrate the computer into assessment in ways which have no paper parallel.

The focus of the study was placed on the FCI for several reasons. First, the FCI is already in wide use. That means that it is known by the community and that there is a base of thought and research data from many places and people dealing with the FCI. It has also made data collection easier since it is easier to convince an instructor to give a test based on one that he or she already knows and uses. The FCI also helped streamline this project because it enabled me to focus on the research questions instead of developing new test questions along with animated versions.

Secondly, the FCI is based on common but *surprising* misconceptions. It seems strange that someone who has interacted with the world for 18+ years could not correctly predict the path of a hockey puck after it was kicked. But in reality, many of those students, who easily and effortlessly navigate their way through the physical world, can not correctly predict the path of the puck. To the expert, the wrong choices for the path look odd on paper, but they look ridiculous in motion. Will the students also see the "silliness" of the wrong responses? Or will their misconceptions hold even on the supposedly more intuitive level of animation?

Finally, the FCI deals with conceptual understanding, an important and debated topic. Most physics teachers agree that conceptual understanding is important but there is not a community-wide agreement about what conceptual understanding is and how (or if) it should be measured. The FCI is given widely every year, and the results are felt to have some meaning, but it is not clear what FCI scores tell us. I have seen instructors give the FCI rigorously every year and interpret the scores as a sign of their students' "conceptual understanding" and their instructional competence. I have also encountered a number of instructors who bitterly detest the FCI for various reasons including the feeling that it "tricks" the students and that the results have no meaning.

Since the area of conceptual assessment in introductory physics is so important and yet so debatable, there is a need for research to shed light on some of the issues. Conceptual understanding has been thoroughly probed on a large scale with pencil and paper. Computer technologies offer a way to assess this understanding from a different vantagepoint which has the potential to inform us about students' understanding and about our abilities to assess that understanding.

And finally, why focus on animations? Well, the use of animations for instruction is not well understood and the use of animations for assessment has almost completely been ignored. Animation provides a way of assessing students that is not possible with pencil and paper. It is possible that they will provide a better way to find out what students know and understand. We are no longer limited to asking questions on paper. It is time to investigate others ways to assess student learning. Even if animations do not provide a better way to assess, the process of making that determination will certainly lead to some interesting insights about the way students think and answer conceptual questions.

---

[1] Rieber, L. (1994). Computers, Graphics, & Learning. Dubuque, IA: Wm. C. Brown Communications, Inc.

[2] Wenglinsky, H. (1998). Does it compute: The relations between educational technology and student achievement in mathematics (Report available from the Educational Testing Service).

[3] Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. The Physics Teacher, 30, 141-158.

[4] http://www.webassign.net/manual/

Chapter 2
Review of the Literature

There are three main areas relevant to the current study. First, the conceptual test that provided the basis for the study will be reviewed. Secondly, a review of the research involving the use of graphics and animations in both educational and psychological contexts will be summarized. Finally, the tool used to create the computer animations will be discussed.

2.1 The Force Concept Inventory

2.1.1 Purpose

The Force Concept Inventory is a conceptual exam, designed to measure students' understanding some of the basic concepts of Newtonian mechanics and to identify which common misconceptions students exhibit. Its primary use is as an assessment instrument. The FCI is generally given as a pre-test at the beginning of a course and then again as a post-test at the end of the course, or after instruction on forces is complete. The results can then be used to evaluate achievement of the class as a whole. As will be discussed, it probably can not give a meaningful analysis of individual students nor is it useful except when scores on the entire test are considered.

2.1.2 History

In 1985, Halloun and Hestenes published the Mechanics Diagnostic Test.[1] According to the authors, this 34-item test was designed to "assess the knowledge state of beginning physics students". The test questions were 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". This statement is slightly misleading because, in large part, the test was designed around misconceptions.

The authors did not design a test covering the important principles, but rather, a test that asks all the questions students tend to get wrong. Because of the design of the test, scores would be expected to be low.

Initially, questions were written and given in open-answer form to numerous introductory level college students. Based on the answers students gave, a multiple-choice version was created which contained the most common open-ended answers as choices. The multiple-choice version was then used to establish validity and reliability.

The Standards for Educational and Psychological Testing[2] define validity as the "appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores." In essence, a test is considered valid if it measures what it intends to measure. In this case, the test should measure students' understanding of motion and forces.

Face validity is defined as "the extent to which a casual, subjective inspection of a test's items indicates that they cover the content that the test is claimed to measure".[3] Face validity on the MDT was established by expert (physics professors and graduate students) opinion and by interviewing students to ensure that they understood the questions and the answer choices. The test only included questions associated with common misconceptions, which was intentional. So content validity was not an issue. A correlation of about 0.5 was found between the pre-test MDT score and course exam average. This establishes predictive validity.

A test is considered reliable if it produces the same results every time it is given under the same conditions. For example, if a student takes the test on a Friday, his or her score should be more or less the same as if the test was taken on a Wednesday, all other factors being equal. Reliability for the MDT was established in two ways. First, students were interviewed about their answer choices. The students gave the same answers in their interviews as they had on the written test, and it was found that the students generally had reasons for their choices. Second, a statistical analysis was conducted. The authors ran a Kuder-Richardson Test (discussed in section 3.6.1) and found a KR reliability coefficient of 0.86 for the pre-test and 0.89 for the posttest. Essentially, the KR reliability coefficient gives an average correlation between separate items on a test. These values indicate that the test is highly reliable for both group and individual use. The high reliability of the pre-test indicates that students come into the class with defined misconceptions since random guessing would have lowered the reliability coefficient.

The authors gave their test to a large number of students at their university and to some local high school students. Two of their more interesting findings from this effort are listed below.

1. Students start out with low scores and improve very little on the MDT after conventional coursework.
2. There is a significant correlation between pre-test score and final course grade. There is an even stronger correlation between post-test score and final course grade.

The 29 item Force Concept Inventory was first published in 1992.[4] It was the result of "improvements" to the MDT. The authors did not elaborate on the nature or reasons of the improvements beyond the somewhat cryptic statement "the Inventory has the advantage of supplying a more systematic and complete profile of the various misconceptions". Apparently, the authors made it a point to cover all areas of the concept of force more completely by adding some questions and eliminating others. However, it should be noted that they still only focused on those areas where common misconceptions are found. Presumably the authors also incorporated any criticisms they had received about the MDT. About half of the questions on the FCI were the same as those found on the MDT, and the purpose was the same.

For purposes of analysis the authors presented a break down of the questions on the FCI into the six conceptual dimensions of kinematics, first law, second law, third law, superposition principle, and kinds of force. (See Table 2.1) They also provided a list of 30 student misconceptions the test probed and which questions addressed each misconception. (See Table 2.2)

Table 2.1 - Newtonian Concepts in the Force Concept Inventory (Version One), taken from The Force Concept Inventory (The Physics Teacher, 30, pp. 141-158).[4]

	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)
1. First Law	
with no force	4B, (6B), l0B
velocity direction constant	26B
speed constant	8A, 27A
with cancelling forces	18B, 28C
2. Second Law	
Impulsive force	(6B), (7E)
Constant force implies constant acceleration	24B, 25B
3. Third Law	
for impulsive forces	2E, 11E
for continuous forces	13A, 14A
4. Superposition Principle	
Vector sum	19B
Cancelling forces	(9D), 18B, 28C
5. Kinds of 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, 18B, 22D
acceleration independent of weight	1C, 3A
parabolic trajectory	16B, 23D

Table 2.2 - A Taxonomy of Misconceptions Probed by the Force Concept Inventory (Version One). Presence of the misconception is suggested by selection of the corresponding Inventory Item. Table taken from The Force Concept Inventory (The Physics Teacher, 30, pp. 141-158).[4]

	Inventory Item
0. Kinematics	
K1. position-velocity undiscriminated	20B,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,B,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	
AF1. only active agents exert forces	11B; 12B; 13D; 14D; 15A,B; 18D; 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	18A,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; 10C,D,E
Ob. Obstacles exert no force	2C; 9A,B; 12A; 13E; 14E
Resistance	
R1. mass makes things stop	29A,B; 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

The authors of the FCI did not establish test validity and reliability for the FCI. They claim that "Considerable care was taken to establish the validity and reliability of the Diagnostic. 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." In justifying why validity and reliability was not re-established the authors pointed out that similar scores were found on both tests in similar populations. So, on the one hand the authors say the test is a different, improved test. And on the other, they claim it is so similar that there is no need to establish validity and reliability.

Only half of the questions from the FCI are the same as those found on the MDT, so it is a different test and reliability and validity should have been reestablished. Although no evidence has since been presented that the test is invalid or unreliable, the lack of established reliability and validity for the FCI at publication time is a shortcoming.

The authors of the FCI did interview a small number of students about their responses to particular FCI questions and found responses to be similar to those found on the MDT and relatively free of false positives. Also, since its introduction, similar pre and posttest scores have been found across the country, when instruction style was similar, which does indicate that the authors were correct in assuming reliability. Also, since its introduction, the numerous physics instructors who have given the test have generally agreed that the test does seem to measure students' concepts of forces.

A new version of the FCI was distributed on the web and at conferences around 1996. It has never been formally published except in Mazur's book on peer instruction.[5] No documentation regarding issues of test analysis has been made publicly available. The new version is meant to be an improvement of the original FCI. Again, the authors do not make it clear why the improvements were made but the changes were probably based on criticisms from the physics education community. A few questions were added, a few questions that were often misunderstood were dropped, and almost every remaining question was slightly reworded to improve readability. It is unclear if the changes in the test were research based.

Again, validity and reliability were not reestablished. Also, no formal documentation has been provided for the test. The tables breaking down the questions and responses by concept are missing. There is no report of the relationship of scores between version one and version two, nor are comparison data publicly available. Over time, other researchers may fill in these gaps. There is a need for more research relating to the FCI to be published.

2.1.3 Research Findings

Since the Force Concept Inventory was first published in 1992, it has been used as a yardstick of instruction in many physics classrooms across the country. At the winter 2000 meeting of AAPT[6], there were at least 10 papers presented that contained FCI in the title or abstract. However, surprisingly few studies looking at the test itself have been published.

In the original paper containing the FCI, the authors reported on some of their own research on FCI results. Their main findings are listed below.

1. Math background is not a major factor in high school FCI scores.
2. Pre-test scores are uniformly low for beginning students.
3. Large gains from pre to post test were not seen with conventional instruction.
4. Even graduate students do not get perfect scores on the FCI.
5. There is a conceptual threshold near 60% on the FCI. Below 60% "a student's grasp of Newtonian concepts is insufficient for effective problem solving."

In general, these results were presented without clear documentation. However, it does appear that the authors took extensive data on the FCI. They published their results in *The Physics Teacher*, which is geared toward practitioners rather than researchers. It is likely that this led to the lack of documentation a researcher would be interested in.

The most notable research result separate from the FCI authors came from Richard Hake.[7] Hake collected data from 62 introductory level courses that administered the

FCI or MDT as a pre-test and post-test. His purpose was to compare scores on the FCI and MDT with type of instruction.

In analyzing the data he collected he proposed what has since been referred to as "Hake's Gain Score" which is defined as

$$g = \frac{(x_f - x_o)}{(100 - x_o)} = \frac{\text{Total Gain}}{\text{Possible Gain}}$$

Where x_f = The post-test class average reported as a percent.
x_o = The pre-test class average reported as a percent.

The gain score provides a way to compare classes that started with very different FCI scores since it normalizes the scores based on how well the class performed before instruction.

Hake computed gain scores for each course and asked the instructors to label their courses as either traditional or interactive engagement. Interactive engagement was defined as methods "designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities with peers and/or instructors". Traditional was defined as methods that "make little or no use of interactive engagement methods, relying primarily on passive-student lectures, recipe labs, and algorithmic-problem exams".

He then looked to see which courses had the highest gains. His results were clear. (See Figure 2.1) The highest gains were associated with courses that were labeled interactive engagement. In fact, all 14 traditional courses had gain scores of less than 0.3. Only 7 of the 48 interactive engagement courses fell in this range.

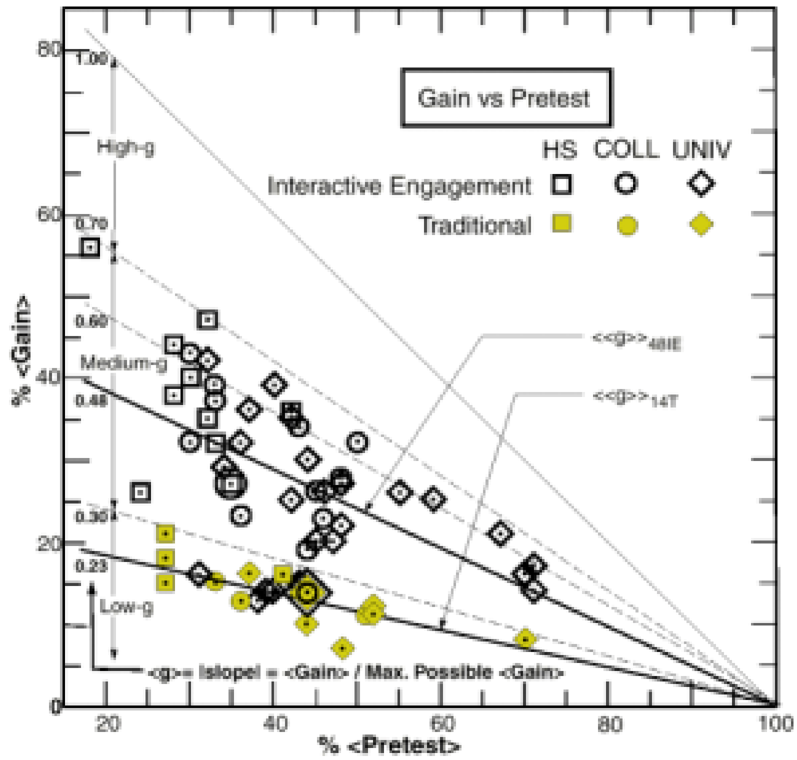


Figure 2.1 %<Gain> vs. %<Pre-test> score on the conceptual Mechanics Diagnostic (MD) or Force Concept Inventory (FCI) test for 62 courses enrolling a total N=6542 students: 14 traditional (T) courses (N=2084) which made little or no use of interactive engagement (IE) methods, and 48 IE courses (N=4458) which made considerable use of IE methods.

The FCI appears to be a good measure of the level of interactive engagement in mechanics based physics courses as judged by instructors. Since it is generally believed that the FCI is measuring student understanding, the publication of Hake's results was extremely important. The established relationship between increases in FCI scores and interactive instructional style has led to the FCI being considered a valid measure of curriculum and instruction by most researchers in physics education.

The only other published study that directly investigates the FCI came from Steinberg and Sabella at the University of Maryland.[8] They investigated how student performance on the FCI correlates with performance on similar exam problems by giving students open ended, non-numerical exam questions similar to some of the FCI questions. The same students who took the exam also took the FCI.

The results of this simple study were mixed. In some cases students performed better on the FCI questions and in some cases, students performed better on the exam question. The researchers made some suggestions about why the results were mixed but they were unable to offer any supporting evidence for their ideas. The variable correlation between FCI questions and similar questions does indicate that student responses on individual FCI questions may not be as reliable as the test taken as a whole.

An unpublished (as of this writing) study has been conducted to determine if using the World Wide Web instead of pencil and paper has an effect on student responses to test questions.[9] The researchers gave the FCI to 197 students and the Maryland Physics Expectation Survey to 204 students. Students were randomly assigned to a group that took the test over the World Wide Web or to a group that took the test on paper. No differences were found between the two groups. Their results indicate that using the web to deliver pencil and paper type diagnostic tests will produce the same results as if the tests were given in their traditional format.

2.1.4 Criticism

Although the FCI has been used extensively and discussion of FCI gain scores is a staple of conversation among physics educators, it is not free of criticism.

After the first version of the FCI was published, the authors received criticism about the wording of many of the questions. For example, a letter published in *The Physics Teacher* complains about two of the questions.[10] The authors demonstrated they took note of these criticisms when they released the new version. Most questions were changed slightly from version to version, several questions were completely removed, and several were added.

In 1995 Huffman and Heller[11] published an article which set off an intense debate about the FCI. In this first article, they discussed the results of a factor analysis they had performed on FCI scores at a university and at a high school. Factor analysis is a complex statistical analysis that involves calculating the correlation between test items and then grouping items together in "factors" that account for the variance. Factor analysis is performed to see how items on a test are related. Questions that measure the same concept should fall under the same factor.

Huffman and Heller found that FCI responses from high school students gave only two factors and responses from university students gave only one factor. The authors of the FCI had proposed that the FCI was made up of, and could be analyzed with, six conceptual dimensions which is inconsistent with the factor analysis findings of Huffman and Heller. In conclusion, Huffman and Heller stated

> The items on the inventory appear to be only loosely related to each other, and instructors should be cautious about concluding that the inventory actually measures students' understanding of a "force concept". It seems more likely that the inventory actually measures bits and pieces of students' knowledge that do not necessarily form a coherent force concept. Caution is also advised in analyzing any of the six conceptual dimensions separately.

Eight months later, Hestenes and Halloun responded to Huffman and Heller.[12] They contend that the clustering of student responses on only one or two factors is not a problem for a number of reasons. They claim that the six conceptual dimensions they proposed were never meant to describe student concepts but to provide a standard to which the student concepts could be compared. They also state that "the FCI should be administered and interpreted as a whole; separate pieces of it are much less reliable and informative." Hestenes' and Halloun's response was reasonable, but they failed to acknowledge that their stance was an adapted version of their original thinking. The test has the most meaning when looked at as a whole. They did not make this clear in their original paper.

In response to the question, "What does the FCI score tell us?" Hestenes and Halloun offer the following. An FCI score of 60% is an entry threshold to Newtonian physics. Students at this level are just beginning to use Newtonian concepts coherently in their reasoning. Students below this level will tend to use Newtonian concepts inconsistently. The authors also suggest 85% as the Newtonian mastery threshold. At this level students are confirmed Newtonian thinkers. They give no concrete evidence to support the 60% and 85% cutoffs. They appear to be arbitrary and intuitive.

The debate continued. In the same issue of *The Physics Teacher*, Heller and Huffman[13] responded to Hestenes and Halloun's response. Much of their second paper was devoted to defending their statistical analysis. They also questioned Hestenes and Halloun's proposal of the 60% and 85% threshold levels by presenting data which shows that correlations between questions is no higher for high scoring students than for students below the mastery threshold. The article concludes with the statement, "more research is needed to determine what the FCI is actually measuring."

The debate was intense, but in the end, they all seem to be making the same essential points. First, student responses on the FCI are not consistent from question to question. Even for those questions that a physicist would consider identical, student responses are not correlated. This is consistent with the findings of other researchers that students' alternative beliefs about physics are ill defined, loosely organized, and dependent on the specific circumstances of the question.[14, 15] Secondly, and most importantly, analysis of the FCI by individual student or individual question is unlikely to yield meaningful results. Hake's data showed that the FCI, taken as a whole does have some meaning while Heller and Huffman have shown that its usefulness dissipates when the whole is broken into pieces.

2.1.5 Importance

Although the FCI has its flaws, it has been extremely influential in the field of physics education. One of its greatest impacts has been to build a bridge between the physics education research community and traditional physics instructors. The questions are simple. Most physics instructors look at them and think that their students should be able to answer them without difficulty. When their students have performed poorly on the FCI, many instructors have begun to rethink their teaching technique. Eric Mazur is the most well known example. He discusses his "ah-ha" moment in his book Peer Instruction.[5]

> "I taught a conventional course consisting of lectures enlivened by classroom demonstrations. I was generally satisfied with my teaching - my students did well on what I considered difficult problems, and the evaluations I received from them were very positive. As far as I knew, there were not many problems in *my* class…The first warning came when I gave the Halloun and Hestenes test to my class and a student asked, "Professor Mazur, how should I answer these questions? According to what you taught us, or by the way I *think* about these things?" Despite this warning, the results of the test came as a shock: The students fared hardly better on the Halloun and Hestenes test than on their midterm examination. Yet the Halloun and Hestenes test is *simple*, whereas the material covered by the examination (rotational dynamics, moments of inertia) is of far greater difficulty or so I thought."

The FCI seems to speak to the traditionalist in a way that the perceived radicals in the field of education can not.

The FCI has also become a standard of curriculum assessment, as discussed previously. For better or worse, it is used extensively in physics classrooms. It is so well known and accepted that FCI scores are discussed at national meetings without reference to what the test is or where it came from. Although it has been occasionally replaced in recent years by the similar Force and Motion Conceptual Exam[16], it will undoubtedly continue to be used in numerous classrooms in the future. Therefore, it is important to continue assessment of the test itself.

2.2 Research on Computer Animations in Instruction

It is only recently that computer technologies have developed to the point where instructors can easily incorporate computer animations and simulations into their instruction at low cost. Twenty years ago the personal computer was just beginning to make its presence felt in the classroom. Even as recently as ten years ago, such features as the mouse and color monitors were fairly new additions to the personal computing environment. Because the ability to easily incorporate computer animations into instruction is new, and because of its rapidly developing nature, there is too little research on how students are affected by and interact with such technology.

In theory, animation should help students to process information for several reasons. First, motion can capture the attention of the learner and help him or her focus on important aspects of the presentation. For example, Brasell[17] found that real-time graphs could help students improve their understanding of distance and velocity graphs. She suggested that the movement of the display "may encourage students to selectively attend to salient points on the graph, such as where there are changes in the physical event (e.g., changes in speed or direction), or motivate them to find out how to create changes in the graph." Of course, the high level of attention associated with motion could also be a major distracter of learning if the motion is not well suited for the learning task.

Graphics can also assist memory. Many studies have found that pictures are remembered better than words in recognition tests.[18] Also, concrete words which can be visualized (e.g. dog, table, tree) are better remembered than abstract words (e.g. knowledge, love, belief) which can not be visualized.[19] It is likely that graphics aid in memory because pictures are easily stored in long-term memory and contain a large amount of information.

2.2.1 Rieber's Review of Research

At the forefront of research into instructional use of computer animations is Lloyd Rieber of the University of Georgia. In addition to contributing a large percentage of the research base he has also written two extensive reviews of the research, one in 1990[20] and another in 1994[21]. Because of his extensive experience in the field his reviews and thoughts are well worth summarizing.

Static Visuals

Static graphics have been available to educators for a long time and considerable research has been done on their effectiveness. Based on his reviews of several reviews of the research in the area of pictures, Rieber offers the following instructional design principles.

- Since the available research is largely based on the effects of pictures on learning from prose, this research and any principles derived from it can only be consistently applied to similar situations.

Most of the research, done on the effectiveness of pictures for learning, concentrates on helping students learn textual information presented as a narrative. However, many instructional goals differ from this emphasis. Although the research base may be applicable to a particular use of pictures, it may not. For non-prose applications of pictures, such as free-body diagrams or concept maps, the instructional designer must decide if these principles can be generalized to the particular task.

- Pictures should be congruent and relevant to the information presented in the text. If pictures are not congruent or relevant to the text, they may be distractive and interfere with learning.

The highest learning gains have come when pictures acted as visual mnemonics. In contrast, pictures that are not related to the text can be harmful to learning. A decorative picture could distract the student from the important information and interfere with the transfer of knowledge into short-term memory. Also, unnecessary graphics could mislead the student into thinking they are important and cause misconceptions.

- Pictures will not be necessary when the text already prompts the learner to spontaneously form internal images.

In cases where the text is vivid enough that students can use it to create their own images then an external image is unnecessary. This depends on the context of the information as well as the students' abilities and background. There is also the ceiling effect. If text alone can lead to almost perfect understanding then there is little room for the additional benefit of a graphic.

- When pictures are prepared as an aid to the learning from text, be sure that the learner can read and understand the text.

This is the opposite of the ceiling effect. If learners are not able to get anything from the text, a picture (which is meant to aid in understanding the text) probably will not be helpful.

- Pictures should be prepared to clearly represent the content that is to be remembered; additional "dressing up" of the picture should be avoided.

In other words, keep it simple. Too much information will make it more difficult for the learner to focus on the relevant parts of the graphic.

Animated Visuals

Rieber suggests that two conditions exist for an animation to offer the potential for increased learning. First, "a need for external aids to visualization must be established … if students adequately image internally, then, obviously, the inclusion of external visuals will probably not result in any additional learning gains." Second, "the learning of the content must depend on understanding either changes to an object over time (i.e., motion) or changes in the direction in which the object is moving (i.e., trajectory), or both." In other words, animation will be most beneficial when the concept to be learned involves understanding changes in speed or direction of motion of an object.

After a review of the research literature, Rieber made three recommendations.

> Recommendation 1: Animation should be incorporated only when its attributes are congruent to the learning task.

In his review of 17 research studies, Rieber reported that about half of the studies found no significant difference between the animated and the static or text only group. (See Table 2.3) Other researchers have confirmed this finding. In a 1996 review of the literature, Szabo and Poohkay[22] found 20 studies investigating animation in the context of computer based instruction. Ten of the studies showed significant effects for the animation and ten showed no significant differences.

Table 2.3 - Summary of Empirical Research on the Instructional Effects of Animation, taken directly from Computers, Graphics, & Learning by Rieber (1994).[21]

Study	**Subjects** *No.*	*Age*	**Content**	**Learning Outcome**	**Results Regarding Animation**
Baek & Layne, 1988	119	High School	Mathematics (ave. speed)	rules	Anim>Static> Text
Comments: Students' attention was focused on information contained in the animation.					
Caraballo, A., 1985	109	Adult	Computation of area of geometric shapes	facts, concepts, rules	NSD
Comments: Animation was used as an aid to conceptual understanding, not as an elaboration of the lesson content.					
Caraballo, J., 1985	109	Adult	Physiology of human heart	facts, compre-hension	NSD
Comments: No pilot studies conducted to determine a need for external visualization.					
Collins, Adams, & Pew, 1978	18	Adult	Geography of South America	facts	Interactive Map> Labeled Map> Unlabeled Map
Comments: Animation for attention-gaining within an interactive graphic (blinking dots).					
King, 1975	45	Adult	Mathematics (sine ratio)	rules	NSD
Comments: Possible confounding due to an easy learning task, verbally heavy tests which may not have been sensitive toward visualization tasks, and crude graphics.					
Mayer & Anderson, 1991	102	Adult	How a bicycle pump works	problem solving	Animation with narration> Animation only= Narration only = Control
Comments: Study involved three separate experiments testing various predictions of dual coding theory. Summary above shows general results of the final experiment. Results of the first two experiments showed that animation *with* narration supports dual coding more than narrations *before* animation.					
Moore, Nawrocki, & Simutis, 1979	90	Adult	Psychophysiology of audition	facts, rules	NSD
Comments: The study was confounded due to instructional design. A review was given to all students after cad of tour lesson parts it they did not achieve at least 85% master on the respective part. This induced an artificial ceiling effect.					
Reed, 1985	180	Adult	Algebra word problems	rules, problem solving	Mixed according to word problem type
Comments: Lessons were iteratively improved over the course of four separate experiments. Animation was effective when replacing rather than supplementing verbal information.					

Table 2.3 - Continued

Study	**Subjects** *No.*	*Age*	**Content**	**Learning Outcome**	**Results Regarding Animation**
Rieber & Hannatin, 1988	111	Children (elem.)	Newton's laws of motion	rules, problem solving	NSD
Comments: There was not enough variance in the lesson treatments. Animation not a powerful enough effect, especially within an orienting activity. Students also found the material very difficult to understand.					
Rieber, 1989	192	Children (elem.)	Newton's laws of motion	facts, rules	NSD
Comments: The study may have been confounded due to the difficulty of the content. Students may not have been attending to the animation appropriately as evidenced by latency data of the time spent by students processing each frame of instruction.					
Rieber, 1 990b	119	Children (elem.)	Newton's laws of motion	rules, problem solving	Animation>Static= None; Interaction between visuals & practice
Comments: Animation most effective as a presentation strategy when practice support was moderate. The structured simulation was effective as a practice strategy.					
Rieber, 1991a	39	Children	Newton's laws of motion	rules	"Chunked" Anim.>Static
Comments: Animation was an effective presentation strategy, but only when screens were presented in parts, or "chunks," to aid students in selectively attending to information in the animated visual.					
Rieber, 1991b	70	Children (elem.)	Newton's laws of motion	rules, problem solving (intentional and incidental)	Animation>Static
Comments: Evidence that the visually based structured simulation was intrinsically motivating for students. Students also learned a rule incidentally from an animated display, but they also became prone to a scientific misconception.					
Rieber, Boyce, & Alkindi, 1991	127	Adult	Acceleration & Velocity	rules (near and far transfer)	NSD on orienting activity; Simulation results mixed
Comments: A visually based simulation was ineffective as an orienting activity, but effective as a practice strategy for near transfer tasks only. Feedback from subjects indicated that they found the content quite demanding. They also seemed uncomfortable with the simulation in that they seemed to expect more structure.					
Rieber, Boyce, & Assad, 1990	141	Adult	Newton's laws of motion	rules, problem solving	NSD on learning; Animation>Static> None on response latency
Comments: Although no differences were found on performance measures, animated presentations may have aided organization and retrieval as evidenced by latency data on posttest. Structured simulation generally effective as practice strategy.					

Table 2.3 - Continued

Study	**Subjects**		**Content**	**Learning Outcome**	**Results Regarding Animation**
	No.	*Age*			
Rieber & Parmley, 1992	160	Adult	Newton's laws of motion	rules	Structured Sim= All Tutorial Groups> Unstructured Sim & Test Only Groups
Comments: Subjects were able to inductively learn from a structured simulation, but not an unstructured simulation. Subjects' response confidence lower without access to traditional tutorial.					
Rigney & Lutz [Alesandrini], 1975	40	Adult	Science: How a battery works	facts, concepts, rules	Pictorial group> verbal group
Comments: Since there was no control for the use of static versus animated graphics, effectiveness of animation cannot be inferred from the results.					

Note: NSD—No Significant Differences

These inconsistent results are an indication that animation in and of itself is unlikely to produce consistent results. Rieber's first recommendation was based on his analysis of the differences between those studies that produced significant results and those that did not. He found that in the cases where the animation did not produce an effect on student learning, the task did not need external visualization, was simple enough that aids were not necessary, was too difficult for an effect to be noticed, or was not supported in the context of a full lesson with practice.

To an extent, his argument is circular. When a study produced no significant results he looked for an excuse, when significant results were found, he assumed the conditions had been met. What is apparent is that animation will not always benefit learning. Whether or not animation is a valuable addition to the learning task will depend on more specific attributes of the animation such as how it is written, and the content to be learned. Just as research on the effectiveness of "books" as an entity would not yield meaningful results, animation needs to be considered in more specific ways.

> Recommendation 2: Evidence suggests that when learners are novices in the content area, they may not know how to attend to relevant cues or details provided by animation.

Beginners in a content area may overlook information contained in an animation even though the information may be obvious to an expert in the field. Without direction to guide them as they view an animation, students may not attend to the relevant details. Several studies have found that the effect of animation is stronger when the animation and relevant supporting materials, such as text, is presented together. This indicates that animation should be structured to help learners attend to the relevant information.

For example, Mayer and Anderson[23] found that students performed better on a transfer problem when they saw the animation at the same time as narration. When students saw the animation either before or after the narration, they performed no better than a group that had no instruction. However, the students in all groups performed equally better than the no instruction control group on a retention test. The animation helped students to develop a better understanding, but only when presented concurrently with narration.

In a very different study,[24] students were asked to play 11 "games" in a computer microworld designed to address misconceptions about Newton's laws of motion. In each game the students had to pilot a spaceship to a set outcome by rotating it and applying thrusts. Although students in the study did perform better on conceptual questions after interacting with the game, the interesting results from the study came from the game tasks the students had difficulty with.

The last game required students to pilot the spaceship, starting from rest, to a target and hit the target at a low speed. Although students had already correctly completed this task in another game, it was in a different context in which the ship also had to turn. Apparently, students focused on the change in direction in the previous task and did not notice the change in speed because they had difficulty with the new situation. The author commented

> …the experimenter was able to notice even small changes in the speed. The apparent discrepancy in perception occurred for two reasons. First, the experimenter knew from her knowledge of the formal physics that the speed should change following the application of an impulse. It was thus perceptible to her because she expected to see it. Many of the students, on the other hand, did not expect to see changes in speed and, so, did not perceive small changes in speed. Second, when the students were focusing on changing the direction of motion, they often did not notice even large changes in the speed of motion. The quality of the feedback is therefore impaired, in this instance, by the students' misconceptions and by their focus of attention.

Although the information was there, and obvious to an expert in the field, it went unnoticed by the students. In this case the students needed an aid to help them attend to important details. The author suggested introducing aids to help the student such as having the spaceship leave a trace of its path and having a readout of the spaceship's velocity.

> Recommendation 3: Animation's greatest contributions to CBI may lie in interactive graphic applications.

It has been established that students learn more when *actively engaged* than they do when passively taking in information. For example, Hake found that students had higher conceptual gains, as measured by the Force Concept Inventory, when they were part of an interactive learning enviroment.[7] Therefore, we can expect that computer animations will produce the most learning when used to create an interactive environment. Fortunately, computer applications lend themselves to interactive techniques because the computer can adapt to the user based on information the user gives the computer. Such a task is difficult to accomplish with pencil and paper.

2.2.2 Using Computer Animations for Assessment

I have only found one study dealing specifically with using computer animations to assess students. It does not give us much insight, but since it is the only one, it is worth summarizing.

Hale et. al.[25] constructed a 26-item general science problem solving test using computer animations. It is a test intended for eighth grade and up. The test was given to 150 high school and university students. From this data the authors concluded that "the computer-based test is highly similar to paper and pencil tests of the same content in terms of how long it takes to complete it, how good and poor students respond to it, how difficult or easy items are, and how consistently the test measures student performances". Unfortunately, the authors offer absolutely no evidence to support this statement. If a control group was used, no mention of it is made.

Given that only one study seems to have looked at using animations for assessment, and it offers little insight, this is an area in which more research needs to be done. Animation offers a unique way to assess students because of the de-emphasis of verbal skills and the interactive nature. The potential value of this method should be investigated further.

2.2.3 Research on Animations within the Context of Newton's Laws

Although concepts from introductory physics have been used in a number of the research studies published, most focused on the effect of the animation and not on the physics concepts involved. For example, in their study of adults learning Newton's laws, Rieber et. al.[26] report interesting details about the effects of animation, but very little information is given about the actual tutorial the students used. There are only a few studies published which help us understand how animation might combine with specific physics concepts.

Anomalous Trajectories Studies

During the 1980's several psychologists investigated the abilities of adults and children to determine when an object's trajectory was inconsistent with nature. These studies are highly relevant to the current project.

McCloskey and Kohl[27] investigated college (not necessarily physics) students' ability to choose the correct path of a ball when viewing a static diagram of the path and when viewing an animation. Their study was conducted in three parts. In part one, they asked 90 volunteers to consider a ball on a string being rotated in a horizontal plane. The students were asked to identify the correct path of the ball after the string breaks. This question is identical to question 7 from the second version of the Force Concept Inventory, though the answer choices were slightly different. One-third of the subjects were given a paper version of the question and asked to choose the correct response from a set of drawings. One-third viewed an animation of the ball being swung but still saw the answer choices on paper. The remaining third viewed an animation of the ball rotating and were asked to chose the correct response from a set of animations depicting each choice. Although the group that saw the trajectories animated gave the correct answer 73% of the time, the no motion group was correct 60% of the time. This difference was not statistically significant. However, it should be noted that only having 30 students in a group may be the reason a significant difference was not found.

In part two of their study, they repeated the experiment with 72 new volunteers. For this experiment the subjects were asked about the path of a ball after leaving a circular tube. The question was very similar to question 6 from the FCI version 2. In this case there was no difference in the performance of the group that saw the motion and the group that only saw static diagrams.

The method of McCloskey and Kohl's study is similar to a part of the present study. Interestingly, their results are also similar as will be discussed in Chapter 4.

Their final experiment was very clever. The researchers gave 50 new volunteers the following task associated with a puck and the circular table depicted below in Figure 2.2.

> You will start with the puck in the black area at the edge of the table. Then you will try to push it so that it goes through the curve and emerges from the far end without touching the sides of the curve. You may start anywhere in the black area around the edge of the table, but you must release the puck before it reaches the line marking the beginning of the curve.

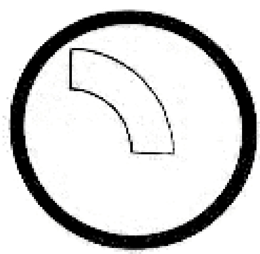


Figure 2.2 - Circular table used in McCloskey study, as depicted in their article.

Interestingly, 67% of the subjects used the correct strategy of pushing the puck straight, 25% attempted to curve the puck before release, in an effort to have it curve after release, and 8% tried to put a spin on the puck so it would curve after release. These percentages are consistent with the percentages of subjects answering in a similar way on the previous two experiments. It should also be noted that men outperformed women on all three tasks.

The researchers had proposed that …

> Most people may have accurate knowledge about how moving objects behave, but this knowledge may be in a form that is not readily activated or can not easily be applied in the context of the rather abstract and static

> problems used in previous studies. Thus, the subjects presented with these sorts of problems may be led to make use of or even to generate theoretical notions (e.g., the curvilinear impetus principle) that are distinct from concrete experience-based knowledge of motion.

This was a very reasonable theory. For experts who do not hold the curvilinear misconception, it is difficult to even imagine that others believe that an object will continue in a curved path in the absence of forces.[5] Student answers to pencil and paper versions of these questions indicate that many students have a misconception. So, as an expert with the correct understanding, it is logical to interpret the student's answer as being an artifact of the testing situation instead of what they really believe.

Although the theory proposed by McCloskey and Kohl was reasonable, the researchers had to abandon it in light of their results. If the students did have accurate intuitive notions about trajectories, they would have been expected to perform better when asked to identify the motion and when asked to recreate the motion themselves. From these results, it must be concluded that the curvilinear impetus misconception is one that some adults do hold. It is not an artifact of the way the question is asked.

Their result is very important in terms of assessment. It indicates the questions are valid when asked in their paper form. Had students answered the questions differently when posed a different way, it would have been an indication that the format of the questions was central to their answers.

Two notes of caution are in order. First, their sample size was small enough that a large difference in performance would have to exist to see a statistical difference. In Experiment One a fairly large, but non-significant difference, was seen. It is possible that the motion did have an effect but the statistical power was too low. Secondly, the animations depicted the ball moving off at a constant speed, even when it took a curved path. Although the correct choice (ball moving in a straight line) should have been at a constant speed, the other choices were curved, and a ball moving in a curved path will not have a constant speed if the trajectory is not circular. When real objects move in a curved path they have a force acting upon them, and will have an acceleration. It is possible that the results would have been different had the speeds been more realistic. However, it is not clear how this would have affected the results as no working theory exists to explain the results.

Two years later Kaiser et. al. essentially repeated phase two of the previous study.[28] In this study, 105 undergraduates were asked to identify the path a ball would take upon leaving a circular tube. However, the question was posed in a rather complex way. There were six possible answers to the question. Observers viewed the possible answers two at a time. Upon seeing each pair, they were asked which of the two seemed to be the most natural. Each subject viewed the 15 pair combinations made from the 6 choices. Half of the subjects saw the pairs in a booklet first and then saw them as animations. For the other half of the sample, the order was reversed. The observers were then classified as either preferring the correct path, an incorrect path, or as being inconsistent.

The researchers then counted the number of observers with a stronger preference for the correct path in the motion or no-motion condition. They found that 47% (significant) of men and 30% (not-significant) of women preferred the correct path in the motion condition but not in the no-motion condition. Performance was not affected by the order of viewing the motion and non-motion condition.

The results of Kaiser et. al. contradict those of the previous study by McCloskey and Kohl. The researchers offer two possibilities for this discrepancy. First, M-K had the observers view all six alternative trajectories in sequence and then pick the best path. This placed a greater memory demand on the observers since they had to remember six possibilities instead of the two in the Kaiser study. Secondly, the observers in the M-K study viewed the animation which ended with a static diagram of the traced path. It is possible that this static diagram played a substantial role. Finally, Kaiser et. al. suggest that since their animations were created using special effect videotapes instead of computer synthesis as M-K's animations, that theirs seemed more natural.

The researchers claimed that their results indicate that "observers are better able to recognize the correct trajectory when presented with the actual events than when viewing static diagrams". While this may be true, it is hard to draw this conclusion from the study. The complex delivery and analysis of the experiment, without articulated reasons for the complexity, make it difficult to determine exactly what they measured. Given this, these results are not highly useful.

The researchers then repeated the experiment with 60 fifth graders. The results were very interesting. Table 2.4 shows the findings.

	Motion	No-Motion
Men	76%	49%
Boys	71%	43%
Girls	63%	31%
Women	32%	23%

Table 2.4 - The percentage of each group preferring the correct path in either the motion or no-motion condition.

The percentages of subjects preferring the correct path in the two conditions are statistically the same for the men, boys, and girls. The women performed at a lower level. This is interesting for two reasons. First, it indicates that males do not improve their views as they interact with the world during their adolescence. This is somewhat discouraging (since 1/4 of male adults have an unexpected misunderstanding of their world) and surprising. Secondly, females actually lose ground. Adolescence seems to move some women from a correct view to an incorrect one. There was no difference in the performance of girls and boys, but as adults, men substantially outperformed women. These findings indicate that the difference seen in adults is a result of women actually shifting from a correct to an incorrect view.

Seven years later Kaiser et. al. report on two other anomalous trajectories studies[29]. In one experiment they asked 48 students to identify the most natural path of a ball, swinging on a string as a pendulum, when the string broke at the highest and the lowest point in the swing. Their method of determining a student's "preferred" path was the same as described in the study above.

Animation improved performance for the highest point of release problem but not for the lowest point of release problem. However, more than 80% of students correctly answered the lowest point problem in the static situation so there was not much room left for improvement, an example of Rieber's ceiling effect. Performance was not affected by the task order (static or dynamic first) and gender effects were noted with men generally outperforming women.

In addition to a static and animated condition the researchers also presented students with what they called a kinematic condition in which students saw the motion, but at a constant speed instead of acceleration appropriate to the motion. They found that improved performance was only associated with motion-with-acceleration and not with the motion-only condition. In other words, when students saw the ball fall from the string and accelerate (as gravity would cause) then they improved their performance over the picture only condition. When they saw the ball fall at a constant speed, the results were the same as the static case. This shows that just the presence of motion is not enough to evoke dynamical intuitions. The motion must mirror reality.

Another experiment, reported in the same paper as the above, dealt with the path of an object dropped from an airplane. The question is similar to the FCI question 14. First, students were asked to view animated trajectories of the fall of the object. The students were then asked to draw the trajectory they had just viewed. Surprisingly, the students often erred. They had a tendency to draw the path relative to the plane rather than relative to the background. For example, when the object fell straight down, 65% drew it as falling backward from the point of release. When it fell forward with half the velocity of the plane 31% drew it as falling backward from the point of release and 14% drew it as falling straight down.

The implication of this result is profound. It appears that incorrect answers may be due more to perceptual issues than to cognitive ones. In some respects, this question is almost an optical illusion for students. If the problem students have in answering this question is perceptual then the validity of this question for conceptual assessment must be considered. This issue will be discussed in the context of the current project in Chapter 4.

Lastly, the researchers compared students' ability to recognize the correct path of fall using the same pair comparison as before. And again, they found animation did improve performance over static pictures.

Identifying Velocities

Zietsman and Hewson[30] report on a study in which 25 tenth-grade students were presented with two objects moving on a rail. The students saw the motions of six pairs of objects and were asked to identify when the objects had the same velocity after each observation. Half of the students viewed the real motions on a physical apparatus. They then saw the motions in a computer simulation. The other half of the students saw the motions in the reverse order. All but 2 students answered consistently in both situations. From these results, it can be concluded that a computer simulation can be a credible representation of reality. Simulations are perceived the same as reality.

Shanon[31] reports on a study involving university students who had not had any course in physics. The study had three phases. In phase one, 25 students responded to two open-ended questions about the time of fall, designed to determine if they thought objects fall at a constant velocity. In phase two, 40 students responded to the same questions but with multiple choice answers. In both phases, a significant number (20-50%) of students answered in a way that was consistent with objects falling at a constant speed.

In phase three, 19 of the students from phase two were shown videos of balls moving at constant velocity and of balls accelerating. The students were asked to describe the motion. All but three of the students classified the accelerating ball as being in free-fall. The response of the other three could not be classified. No students classified the ball moving at constant velocity as being in free-fall. In addition, 35 other students from phase two were also shown videos of balls moving under constant and zero acceleration. This group of students were asked whether is was necessary to suspend the ball from a string in order to produce the motion. They were also asked the same time of fall questions on the written survey. Of the responses that could be classified, 78% were Newtonian.

Although many of the students gave answers on paper that indicated that they believed objects fall with a constant velocity, none of them identified the fall at a constant velocity as being natural. Shanon has identified an area in which people are able to correctly identify natural motion but are not always able to use their knowledge cognitively.

2.2.4 Multimedia-Focused Problems

Titus[32] has investigated student problem solving in physics using video and computer animations. In one phase of his study, he presented video along with a problem statement to about 100 students. Another 100 students saw the same problem statement but without the accompanying video. In each case, viewing of the video was unnecessary because the problems could be solved using the information given in the problem statement. Significant differences were found on three of seven problems using a one-tailed *t* test. However, it should be noted that none of the differences would have been significant had a two-tailed *t* test been used. Although it could be argued that a one-tailed test could be used, since the video was purely an addition and should only help students, this argument does not consider that the addition could be distracting or misleading and could actually cause students to miss a problem they might otherwise solve correctly. As my results will show, using a one-tailed test is not justified here so all results should be considered non-significant.

In another phase of the project, Titus asked students to solve what he called multimedia-focused problems. These problems were computer animations that the students had to interact with in order to get information necessary to solve the problem. These problems proved to be more difficult for the students than their traditional counterparts.

Titus then conducted problem solving interviews with students working on multimedia-focused problems. He observed that,

> 1. A student may make certain measurements before knowing which measurements are needed to solve the problem.
> 2. Viewing the animation and making measurements may cause students to alter loosely held conceptual beliefs.
> 3. A basic understanding of position, velocity, and acceleration is imperative to being able to use measurements to calculate important quantities.
> 4. A student may incorrectly solve a problem because of conceptual difficulties.
> 5. When a student has a conceptual difficulty, he or she may make assumptions which are inconsistent with the problem and animation.

Based on his research, Titus believed that interactive animation may help students to develop a more expert-like approach to problem solving because typical novice approaches are not helpful. The multimedia-focused problems seem to require a greater level of conceptual understanding and level of planning of problem solution. Imbedded in a supportive learning environment, this type of problem could benefit students.

2.2.5 Other Interesting Findings from Animation Research

Animation Preferred by Novices, Text by Experts

Rieber et. al.[33] found that virtually all the students in their study preferred graphical feedback initially when learning the laws of motion. However, textual feedback was generally preferred after the students gained "sufficient experience with the simulation". In the authors' opinion

> …the textual feedback became important and meaningful only after they had an opportunity to build some working hypothesis about the underlying principles based on their interaction with the graphical feedback.

We often communicate complex ideas with a more concise representation. For example, $v = \Delta x/\Delta t$ is a mathematical representation of the concept of average velocity which contains a substantial amount of information. As most physics instructors have discovered, this equation does not convey the same information to the beginning student as it does to his or her instructor. The finding of Rieber et. al. is consistent with this phenomenon. "Simple" representations are only simple once the underlying concepts are understood. Animation has the potential to provide a more intuitive representation.

Animation can Reduce the Time Needed to Process Information

Several studies found that the time required for students to complete a task was less when an animation was used. For example, Rieber et. al.[33] found that when students got animated feedback on the effects of applying a force on an object, instead of textual feedback, they completed the activity in significantly less time. Rieber got the same result in another similar study[34] in which students were learning about velocity and acceleration. Although both the animation and text group performed the same on a post-test, the animation group interacted less with the tutorial and spent less time than the text group.

In another study directed by Rieber [26], it was found that students in the animated graphic condition took significantly less time to answer the post-test than the students in the no graphic or static graphics condition. All groups in this study scored equally on the post-test. The authors concluded that,

> The response latency data indicates that although animation did not affect learning, it helped to decrease the time necessary to retrieve information from long-term memory and then subsequently reconstruct it in short-term memory. This retrieval process required students to reconstruct mental images. The animated visual presentations significantly aided students in this mental visualization effort as compared to students in the static or no visual conditions, presumably by facilitating initial encoding.

These results consistently show that animation may make learning less time-consuming for students even if the end result in performance is the same. This is a not so obvious, but important benefit of carefully designed computer animations.

Animation does not Necessarily Replace the Kinesthetic Experience of Real-Life

Beichner[35] reports on a study in which students were shown video animation of a moving object while a real-time kinematics graph was drawn by the computer. The performance of this group of students on a graphing test was not different than their traditional counterparts. This result was surprising since Brasell[17] had earlier shown that students' performance improved when they generated the motion upon which the graph was based. Apparently, when students are the source of the motion, and can change the motion as the graph is being created, they benefit in a way that can not be replaced by animation.

Different Students May Respond Differently to an Animation

Chanlin reports on a series of studies, which indicate that the effect of an animation may depend on the background of the student. In one study[36] the researcher found that high prior knowledge students benefited from the animation only but that low prior knowledge students benefited from both the animation and static graphics equally. In another study[37], the researcher found that boys benefited from a self-controlled visual presentation whereas girls were just as successful on a post-test when they saw the self-controlled presentation or the system-controlled presentation. These studies indicate that results of instruction using animation with one population should not be applied to all populations.

Chanlin's finding that high prior knowledge students benefited from the animation is somewhat contradictory to Rieber's finding that experts prefer text since students would presumably prefer what benefits them. It should be noted that Chanlin's finding is based on the students' performances and Rieber's is based on the students' preferences. It is possible that experts prefer text even though animation may help them. However, it is most likely that the difference is due to the different learning environments the students were interacting in. This is another indication that the effect of animation will depend on many factors.

2.2.6 Summary of Animation Research

The research that has been reported to date indicates that animations can help students to learn facts and develop concepts. However, it is clear from the research base that the outcome will not always favor animation. A portion of the published studies reported no significant difference between the animated group and the non-animated group. It should be noted that the incidence of no effect animations is probably higher than published research indicates because studies with a null result are less likely to be published.

It is encouraging that no study has found animation to have a negative effect and that the animation often had a positive effect. However, the research literature is lacking in an explanation for why the animation can affect student learning that is based on experimental results. At this point, it is difficult to predict the effect a new animation might have on student learning if previous research is used as a guide. We could make an educated guess, but it would only be a guess. One of the goals of the current research project is to begin to understand *why* animation affects students' ability to answer conceptual questions.

2.3 Physlets

A Physlet® [38, 39] is a **phys**ics app**let** written by Wolfgang Christian, which is scriptable using JavaScript. A number of physlets are available which cover most areas of introductory physics.

2.3.1 Scripting Animator Animations

The Physlet *Animator* was used to create the animated Force Concept Inventory, which is the subject of this dissertation. *Animator* allows the question author to create objects and define their motions. Students can then view the objects' motions in real-time.

Animations are created using *Animator* by embedding code in an html document. Below is a portion of the html and JavaScript code for Question Four of the animated Force Concept Inventory. The following script and explanations are only meant to serve as a general guide. Any reader interested in the details of scripting or of this particular script should consult the Physlet book referenced above.

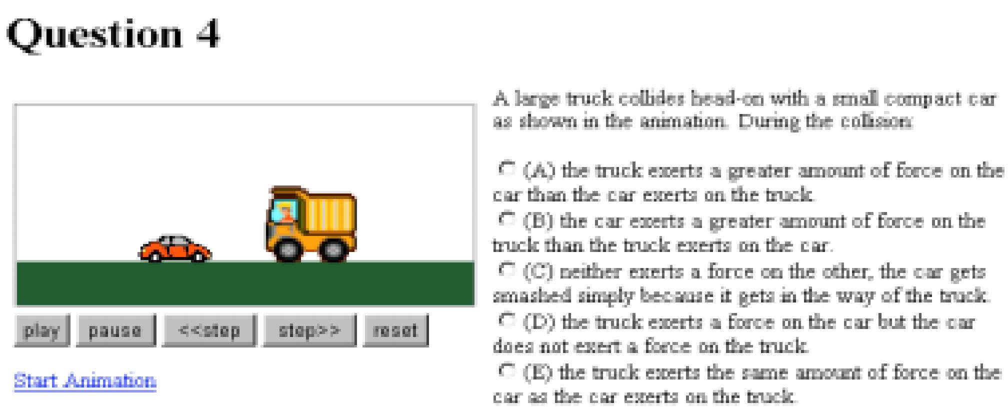
```
<SCRIPT LANGUAGE="JavaScript">
function prob4()
{
      document.Animator4.setDefault();
      document.Animator4.setTimeDisplay(false);
      document.Animator4.setGridUnit(0);
      document.Animator4.shiftPixOrigin(0,-20);
      document.Animator4.setShapeRGB(50,100,50);
      document.Animator4.addRectangle(320,35,"0","-3.8");
      document.Animator4.addImage("smcar.gif","step(0.29-t)*(20*t-
                  11)+step(t-0.3)*(-15*(t-0.3)-5)","0.5");
      document.Animator4.addImage("lgtruck.gif","step(0.29-t)*
                  (-20*t+4)+step(t-0.3)*(-15*(t-0.3)-2)","4");
      document.Animator4.setOneShot(0,1.65,"End of Animation");
      document.Animator4.forward();
}
</SCRIPT>

<APPLET
    code="animator3.Animator.class" codebase="./classes"
    archive="Animator3.jar" id="Animator4" name="Animator4"
    width="300"   height="150" >
<PARAM name="FPS" value="10"> <PARAM name="dt" value="0.05">
<PARAM name="pixPerUnit" value="10"> <PARAM name="gridUnit" value="1">
<PARAM name="showControls" value="false">
</APPLET>

<BR><INPUT TYPE="BUTTON" VALUE="play"
onClick="document.Animator4.forward()">
<INPUT TYPE="BUTTON" VALUE="pause"
onClick="document.Animator4.pause()">
<INPUT TYPE="BUTTON" VALUE="<<step"
onClick="document.Animator4.stepBack()">
<INPUT TYPE="BUTTON" VALUE="step>>"
onClick="document.Animator4.stepForward()">
<INPUT TYPE="BUTTON" VALUE="reset"
onClick="document.Animator4.reset()">
<P>
<A HREF="JavaScript:prob4()">Start Animation</A>
```

The first section is the JavaScript. It sets to default parameters, turns off the time display and drawing of the grid, places the origin off center, sets the drawing color, and draws a rectangle (the road). The next two commands add the images of the car and truck to the applet. The motion is specified in the addImage command as follows: addImage("image source", "x(t)", "y(t)"). A command is then given to play the animation for a specified amount of time and the animation is then set in motion.

The applet section of the code specifies where the code to run the applet is located, and fixes attributes of the applet such as its size and the frames per second. The last section adds buttons to control the playing of the animation.

The code shown above is then embedded in an HTML document. Question text can be added and the various elements can be aligned. The result is shown in Figure 2.3.

Figure 2.3 - Applet from Question 4 of the animated FCI.



2.3.2 Advantages and Disadvantage of Physlets

Advantages

There are a number of advantages in using Physlets for instructional computer animation.

- Physlets can be delivered to students via the World Wide Web. Therefore students can interact with the physlet from any location, at any time as long as they have an Internet connection.
- Physlets are cheap. Physlets are freely distributed for noncommercial use. The only cost of creating an animation using a Physlet is the time necessary to script the animation.

- Scripting an animation is relatively easy and fast once the basic process is understood.
- Physlets are flexible. A physlet is a tool that can be used to create an animation. The animation designer has control over the specifics of the animation.
- Physlets are interactive. Animations can be scripted in a way that requires students to interact with the animation. Also, Physlet animations allow the student to control the viewing of the animation using VCR-like controls.
- Physlets are simple. As previously discussed, research on animations and static graphics has shown that students learn the most from a graphic when it is simple. Too many details can make it difficult for students to focus on the important aspects. It is easy to reduce a situation to its bare bones using Physlets.
- Non-physical animations can be produced. Motion is created through mathematical equations. This allows for false representations of reality to be created almost as easily as accurate representations.

It should be noted that there are other software packages that could be used to create physics animations. The Animator Physlet was used because of all the advantages listed above and because it easily did everything that was required.

Disadvantage of Physlets

- Physlets are slightly ahead of mainstream technology. Physlets use Java 1.1, which is not supported on early versions of the popular browsers (i.e. Netscape $\leq$ 4.05 and Internet Explorer $<$ 4.0). Also, the applets can not be run at all on a Macintosh. Presumably, this problem will disappear over time as technology develops.

Physlets are inexpensive, easy to use, and adaptable to individual classroom environments. For this reason they are being used with increasing frequency. As of this writing, Prentice Hall offers Physlets on their website as a companion to one of their introductory level textbooks, with plans to add two more textbooks to the list of

Physlet supported texts. The Physlet web site gets approximately 5000 hits/month, and there are an estimated 100 regular users.[40] These numbers are expected to increase over time. In addition to their use in instruction, Physlets have been used as a research tool by Aaron Titus[32] and myself, as the subject of this dissertation.

---

[1] Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. American Journal of Physics, 53, 1043-1055.

[2] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

[3] Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction (6th edition). White Plains, NY: Longman.

[4] Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. The Physics Teacher, 30, 141-158.

[5] Mazur, E. (1997). Peer instruction: A user's manual. Upper Saddle River, NJ: Prentice Hall.

[6] AAPT Announcer, 29(4), 1999.

[7] Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66, 64-74.

[8] Steinberg, R. N., & Sabella, M. S. (1997). Performance on multiple-choice diagnostics and complementary exam problems. The Physics Teacher, 35, 150-155.

[9] MacIsaac, D., Cole, D., & Cole, R. (1999, March). Standardized testing in physics via the World Wide Web. Paper presented at the meeting of the National Association of Research in Science Teaching, Boston, MA.

[10] Hudson, R. C., & Munley, F. (1996). Re-score the Force Concept Inventory!. The Physics Teacher, 34, 261.

[11] Huffman, D., & Heller, P. (1995). What does the Force Concept Inventory actually measure?. *The Physics Teacher, 33*, 138-143.

[12] Hestenes, D., & Halloun, I. (1995). Interpreting the Force Concept Inventory: A response to March 1995 critique by Huffman and Heller. The Physics Teacher, 33, 502-506.

[13] Heller, P., & Huffman, D. (1995). Interpreting the Force Concept Inventory: A reply to Hestenes and Halloun. The Physics Teacher, 33, 503-511.

[14] Minstrell, J. (1992). Facets of students' knowledge and relevant instruction. In R. Duit, F. Goldberg, & H. Niedderer (Eds.), Research in Physics Learning: Theoretical Issues and Empirical Studies (pp. 110-128). Kiel, Germany: Institute for Science Education.

[15] DiSessa, A. A. & Sherin, B. L. (1998). What changes in conceptual change? International Journal of Science Education, 20, 1155-1191.

[16] Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula. American Journal of Physics, 66, 338-352.

[17] Brasell, H. (1987). The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. Journal of Research in Science Teaching, 24, 385-395.

[18] Ellis, H. C. & Hunt, R. R. (1993). Fundamentals of Cognitive Psychology. Madison, WI: Brown and Benchmark Publishers.

[19] Driscoll, M. P. (1994). Psychology of Learning for Instruction. Boston: Allyn and Bacon.

[20] Rieber, L. (1990). Animation in computer-based instruction. Educational Technology Research and Development, 38, 77-86.

[21] Rieber, L. (1994). Computers, Graphics, & Learning. Dubuque, IA: Wm. C. Brown Communications, Inc.

[22] Szabo, M. & Poohkay, B. (1996). An experimental study of animation, mathematics, achievement, and attitude toward computer-assisted instruction. Journal of Research on Computing in Education, 38, 390-402.

[23] Mayer, R. E. & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. Journal of Educational Psychology, 84, 444-452.

[24] White, B. Y. (1984). Designing computer games to help physics students understand Newton's laws of motion. Cognition and Instruction, 1, 69-108.

[25] Hale, M. E., Oakey, J. R., Shaw, E. L., & Burns, J. (1985). Using computer animation in science testing. Computers in the Schools, 2, 83-90.

[26] Rieber, L. P., Boyce, M. J., & Assad, C. (1990). The effects of computer animation on adult learning and retrieval tasks. Journal of Computer Based Instruction, 17(2), 46-52.

[27] McCloskey, M. & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 146-156.

[28] Kaiser, M. K., Proffitt, D. R., & Anderson, K. (1985). Judgements of natural and anomalous trajectories in the presence and absence of motion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 795-803.

[29] Kaiser, M. K., Profitt, D. R., Whelan, S. M., & Hecht, H. (1992). Influence of animation on dynamical judgments. Journal of Experimental Psychology: Human Perception and Performance, 18, 669-690.

[30] Zietsman, A. I. & Hewson, P. W. (1986). Effect of instruction using microcomputer simulations and conceptual change strategies on science learning. Journal of Research in Science Teaching, 23, 27-39.

[31] Shanon, B. (1976). Aristotelianism, Newtonianism and the physics of the layman. Perception, 5, 241-243.

[32] Titus, A. P. (1998). Integrating Video and Animation with Physics Problem Solving Exercises on the World Wide Web (Doctoral dissertation, North Carolina State University).

[33] Rieber, L. P., Smith M., Al-Ghafry, S., Strickland, B., Chu, G., & Spahi, F. (1996). The role of meaning in interpreting graphical and textual feedback during a computer-based simulation. Computers and Education, 27, 45-58.

[34] Rieber, L. P. (1996). Animation as feedback in a computer-based simulation: Representation matters. Educational Technology Research and Development, 1, 5-22.

[35] Beichner, R. J. (1990). The effect of simultaneous motion presentation and graph generation in a kinematics lab. Journal of Research in Science Teaching, 27, 803-815.

[36] Chanlin, L. (1999). Visual treatment for different prior knowledge. International Journal of Instructional Media, 26, 213-219.

[37] Chanlin, L. (1999). Gender differences and the need for visual control. International Journal of Instructional Media, 26, 329-335.

[38] Novak, G. M., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). Just-In-Time Teaching: Blending Active Learning with Web Technology. Upper Saddle River, NJ: Prentice Hall.

[39] Christian, W. & Belloni, M. (2000). Physlets. Upper Saddle River, NJ: Prentice Hall.

[40] Christian, W. (2000). Personal communication.

Chapter 3
Quantitative Results

The data for this study was collected in two phases. In the first phase a large amount of numerical data was collected. In this chapter the research and analysis methods used with this numerical data will be discussed along with the results. Although some discussion of the results is in this chapter, more discussion will be found in the next two chapters. The second phase of data collection involves qualitative interview data which will be presented in Chapter 4.

3.1 Research Instrument

All 30 questions from the second version of the Force Concept Inventory were animated using *Animator*. The animations were embedded into an html document so that the Animated FCI (AFCI) could be delivered over the World Wide Web. Appendix A contains a copy of the user side of the AFCI. The traditional FCI is found in Appendix B. Students could submit their answers to the AFCI through the use of an html form. The results were then collected using a Perl-based cgi script.

Every attempt was made to make the animated version of each question as similar as possible to the original paper version, with the obvious exception of the animation. If the original question had a picture then that image was used as the basis of the animation. Since the FCI was not designed as a test to be adapted to animation, accomplishing this conversion was sometimes tricky. Also, the way to go about animating some questions was not obvious. For these reasons, details of the animation of individual questions will be discussed below.

Questions 1, 2, 6, 7, 8, 12, 14, 21, 23, 26, 27

These questions were straightforward to animate. The original version of each question contains a problem statement followed by five answer choices which are a description of motion or the path of a motion. The animated version contains five separate animations which the student must view and choose between. Each animation depicts the motion described in the corresponding traditional answer choice.

For the "What is the path of…" questions there was no information given in the traditional problem statement about speed of the object as it transversed the path. However, it is impossible to animate a path without including speed information. Therefore, I decided to animate the paths according to the following scheme. When an object traveled in a straight path it was given a constant speed, with the exception of 14b where the ball was shown to fall under the acceleration of gravity. When the object traveled in a curved path it had an acceleration consistent with the curve of the path. In general, this meant an acceleration in either the x or y directions and a constant speed in the other direction.

Question 3

This was one of the more difficult to animate, largely because the answer choices are inconsistent. For example, the first answer choice describes a motion but gives no reason for the motion. The last two choices don't describe a motion but suggest reasons. It would be possible for a student to feel both A and D were correct since they deal with different aspects (motion and reason) of the question. Choices B and C have similar problems.

An area was left free at the top of the applet on the animated version of the question. When an answer choice suggested a reason, it appeared in writing in this area. Also, real-time speed was displayed for each choice. For choices A, B, and C the motion was animated as described. For choices D and E the term "falls" was interpreted to mean falls at constant speed and the stone was shown falling with a constant speed from the beginning to the end of the fall.

Questions 4, 5, 11, 13, 15, 16, 17, 18, 25, 28, 29, 30

These questions asked about the forces in a situation. Only one animation was made, which depicted the situation described in the problem statement. The one exception was question 11 which had five corresponding animations for the five choices of path. In the case of question 29, the "animation" was actually a still graphic of a chair at rest.

Questions 9, 10, 22, 24

These questions deal with the speed of an object after an event. In the traditional version, students must reason through the question based on the path alone. However, the animated version had to include speed information so the students could just look at the animation to determine the speed instead of basing their answers on their own ideas. To discourage students from getting speed information directly from the animation, the pause and step controls were removed so that the motion could only be viewed in its entirety, thus eliminating the ability to measure the speed easily. This issue presents a limitation of these questions for comparison purposes. However, it should be noted that in one-on-one interviews with students, no student seemed to use the displayed speed to answer the question.

Question 19, 20

The original versions of these questions show a strobe photo of two blocks moving on a graph. The animated version shows the two blocks moving on the same graph leaving "footprints" at the times corresponding to those shown in the traditional question.

3.2 Data Collected

Student responses to animated and traditional FCI questions were collected in the fall of 1999 from two different sources. In both cases, students answered the questions in class as a non-graded pre-test.

Sample 1

The first sample consisted of 53 students (31 male, 22 female) taking conceptual physics at an affluent private high school located in a suburban area of the South. Students were randomly assigned to one of two groups. Group I took a mixed version of the FCI in which questions 1, 4-6, 8-13, 19, 25-27, and 29 were in animated form and the remaining questions appeared in their original form. Group II took a mixed version opposite to that of Group I. All the questions appeared together on the computer and students submitted their answers electronically. Each student worked at his or her own computer. The mixed versions were used so that each person was part of both the treatment and the control group. This should decrease the overall effect of random fluctuations.

The animated questions were divided between the two groups in such a way that the number of animated questions, and the types of animated questions were about the same for both groups. For example, Question 6 and Question 7 both ask about the trajectory of a ball after a radial force is removed. So Question 6 was animated for group A and Question 7 was animated for group B. Likewise, the puck questions (8-11) were animated for group A while the rocket questions (21-24) were animated for group B. This type of dividing was done with all of the questions.

Sample 2

The second sample consisted of about 325 students (actual numbers varied from question to question) taking calculus-based mechanics for science and engineering majors at a large public university in the Mid-West. Due to a limited number of computers, about 105 students took the fully-animated version of the FCI and submitted their answers electronically. The remaining students took the traditional FCI and submitted their answers on a computer bubble-sheet. Students were randomly assigned to one of the two groups as they entered the laboratory. The students in this sample were mostly male (75%) and white (80%). There was a technical difficulty with the reporting of student answers on the animated version of question 10 so there is no data from this group for that question.

3.3 General Analysis of Data

3.3.1 Did the animation have an effect?

The number of students giving a particular response was compared for the two groups. Comparisons were made for each response (A, B, C, D, and E) for all 30 questions using a z-test for the equality between two proportions. The test statistic is defined as

$$z = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$p_x = \frac{\text{responses in sample x}}{n_x} \qquad P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

Once the z-score is obtained it can be compared with a normal distribution. For these comparisons a two-tailed test was used since it was not known if the animation would have a positive or negative effect. Also, since a large number of tests were performed (5 responses x 30 questions = 150 tests) the level of statistical significance was chosen to be $p = 0.01$ ($z = 2.58$). At this level 1% of the tests would be expected to yield a false positive. In this case that would mean one or two tests. Had a *p*-value of 0.05 been used, which is the norm, then 7 or 8 tests (5% of tests) would have likely yielded a false positive which could obscure trends. This does seem to be the case with this analysis. If all differences at the $p = 0.05$ level are considered general trends are difficult to see, but at the $p = 0.01$ level definite trends emerge.

Table 3.1 lists the results of this analysis by question. The total number of students giving each answer choice is shown for each condition along with the *z*-score used for comparison of the two groups. The significant differences are shaded.

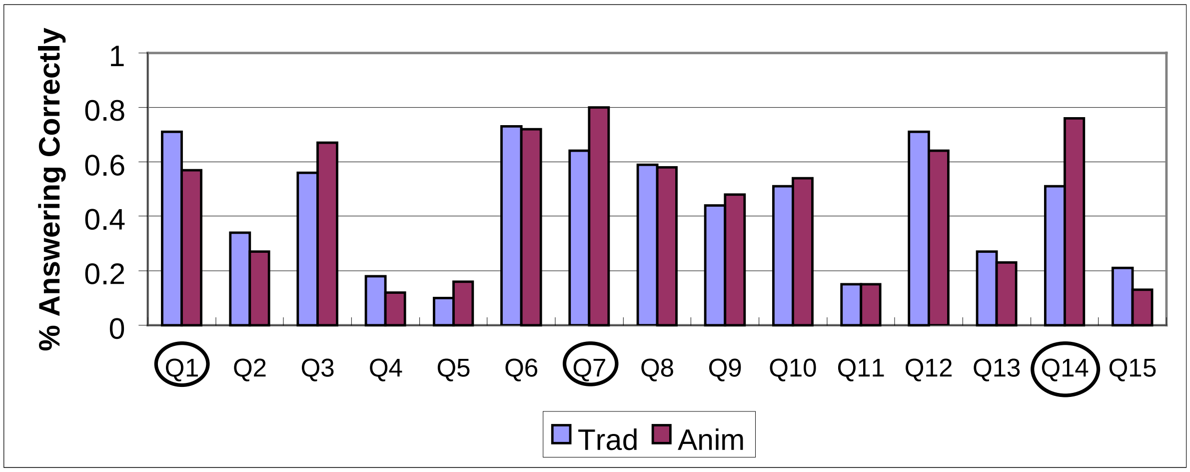
Table 3.1 - The number of students giving a particular response to each question. Except for question 10, the table reflects both sample 1 and sample 2. The *z*-score comparing the number of students giving a particular response in the animated and traditional condition is also given for each question. Significant differences ($p < 0.01$) on a two-tailed *z*-test are shaded and in bold. Bar charts for each question, showing responses by percent, are found in Chapter 4.

		Question														
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
	Response															
Animation Condition	**A**	7	37	10	119	11	31	14	22	6	14	17	0	8	13	18
	B	2	57	30	0	22	97	107	77	25	0	42	85	35	13	7
	C	77	11	91	0	38	5	6	0	27	4	38	41	57	6	102
	D	49	23	2	0	26	0	5	5	12	6	20	7	31	102	7
	E	0	8	3	16	37	1	2	29	64	2	17	0	1	0	0
	Total N	135	136	136	135	134	134	134	133	134	26	134	133	132	134	134
Traditional Condition	**A**	25	89	46	210	13	49	25	31	7	6	36	2	26	74	54
	B	8	51	45	3	26	194	168	154	51	0	83	188	47	32	24
	C	187	16	149	2	71	16	21	3	64	5	84	64	120	23	175
	D	32	92	14	3	69	4	13	18	24	10	39	5	70	133	8
	E	13	15	10	47	85	1	35	57	117	6	21	4	0	0	1
	Total N	265	263	264	265	264	264	262	263	263	27	263	263	263	262	262
Z-score	**A**	-1.48	-1.35	**-2.75**	2.2	1.3	1.08	0.29	1.31	0.96	2.37	-0.28	-1.01	-1.28	**-4.22**	-1.75
	B	-0.93	**4.8**	1.22	-1.24	1.9	-0.23	**3.22**	-0.13	-0.18	0	-0.04	-1.54	2	-0.75	-1.38
	C	**-2.7**	0.76	2.03	-1.01	0.31	-0.98	-1.32	-1.24	-0.94	-0.3	-0.73	1.38	-0.46	-1.55	1.92
	D	**5.7**	**-3.78**	-1.85	-1.24	-1.49	-1.43	-0.56	-1.24	-0.06	-1.11	0.03	1.84	-0.67	**4.86**	1.07
	E	**-2.62**	0.07	-0.85	-1.53	-0.94	0.49	**-3.84**	0.03	0.62	-1.48	1.51	-1.43	1.41	0	-0.72
Correct Response		C	A	C	E	B	B	B	B	E	A	D	B	D	D	A

Table 3.1 - Continued

		Question														
		Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
	Response															
Animation Condition	A	72	97	6	11	6	12	17	13	79	3	49	33	1	24	2
	B	10	6	12	12	5	22	71	46	3	13	13	12	13	48	13
	C	42	0	29	11	83	21	2	23	35	14	19	79	7	0	22
	D	7	24	38	52	31	16	37	28	1	70	18	2	65	51	2
	E	3	4	48	45	6	58	2	20	10	28	27	0	40	2	84
	Total N	134	131	133	131	131	129	129	130	128	128	126	126	126	125	123
Traditional Condition	A	148	190	4	41	34	23	68	26	176	8	91	44	3	49	8
	B	12	22	34	11	12	40	114	103	15	31	97	28	20	101	18
	C	88	7	34	13	68	68	9	50	50	40	18	172	24	5	41
	D	10	28	91	60	120	27	58	65	7	133	33	7	139	81	7
	E	4	14	97	131	23	99	7	11	8	41	12	1	57	5	165
	Total N	262	261	260	256	257	257	256	255	256	253	251	252	243	241	239
Z-score	A	-0.52	0.26	1.77	-2.08	**-2.65**	0.11	**-2.99**	-0.06	-1.38	-0.45	0.5	1.99	-0.39	-0.26	-0.95
	B	1.18	-1.4	-1.18	1.91	-0.39	0.38	1.95	-0.95	-1.54	-0.6	**-5.71**	-0.47	0.67	-0.65	0.98
	C	-0.45	-1.89	2.23	1.28	**7.05**	-2.24	-1.09	-0.45	1.74	-1.29	2.43	-1.08	-1.42	-1.62	0.17
	D	0.65	2.09	-1.28	**3.34**	**-4.4**	0.56	1.29	-0.86	-1.26	0.39	0.3	-0.72	-1.03	1.36	-0.75
	E	0.51	-1.03	-0.24	**-3.14**	-1.55	1.21	-0.73	**3.78**	2.05	1.36	**5.01**	-0.71	1.72	-0.31	-0.14
Correct Response		A	B	B	E	D	E	B	B	A	C	E	C	E	B	C

Ten of the thirty questions showed a significant difference in the distribution of responses between the group that saw the animation and the group that saw the traditional version of the question. Of these, there were three (questions 1, 19, and 20) for which the traditional group performed significantly better and three (questions 7, 14, and 26) for which the animated group performed significantly better. For the remaining four questions where a significant difference was found in the distribution of responses, the animation had no statistically significant effect on the students' ability to answer correctly. In other words, the animation had the effect of shifting students between wrong answers. Table 3.1 shows the differences in correct answers for the two groups.



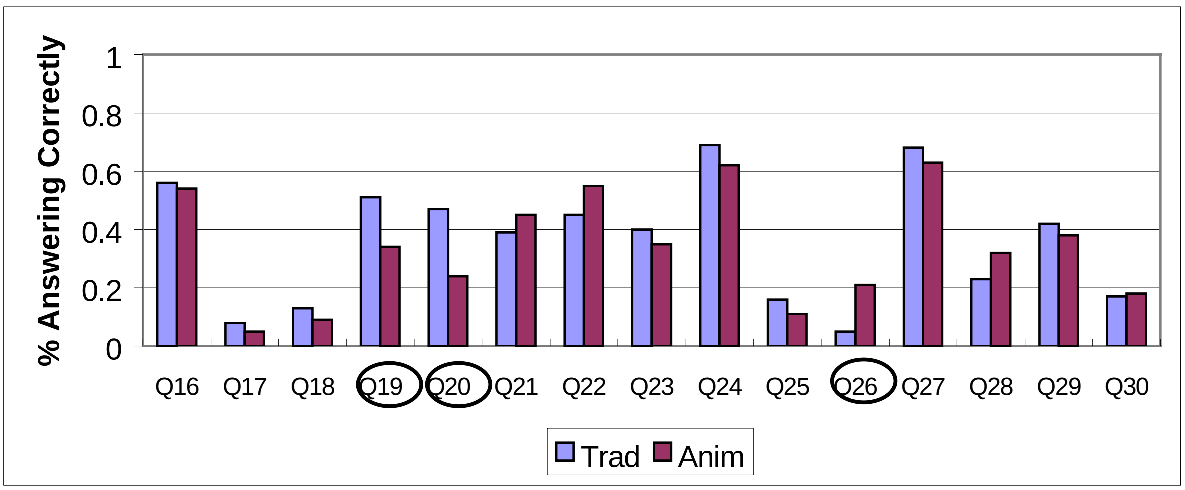


Figure 3.1 - Significant differences in the proportion of students responding correctly. Those questions with significant differences at the $p = 0.01$ level are circled.

There is no question that animation can affect the way students will answer questions like those found on the FCI. However, it should be noted that for the majority of questions (2/3) there was no statistically significant effect. Also, for those questions for which an effect was seen, the animation was sometimes helpful to students and sometimes not helpful.

3.3.2 Can generalizations be drawn about the questions for which an effect (or no effect) was seen?

Topic of Question

I classified the questions using three different schemes. The schemes were derived as the analysis was conducted so they are the result of both pre-conceived categories and data emergent categories. In the first scheme they were classified by the kind of questions asked. Some questions (6-8, 12, 14, 21, and 23) ask for the path (shape of the trajectory) an object would follow. For the remainder of this dissertation these questions will be referred to as "Path Questions". Others ask about the forces on an object (4, 5, 11, 13, 15-18, 25, and 28-30) and will be referred to as "Force Questions". Another group, henceforth known as "Motion Questions", concern the speed or acceleration of an object (1, 2, 9, 10, 19, 20, 22, 24, 26, and 27). At first glance, it is not obvious that questions 1 and 2 belong in the motion group. They were put in this group because it is believed that students generally answer these questions based on the relative speeds of each ball in the x and y directions. This belief was supported in the think-aloud interviews discussed in the next chapter. Question three was considered uncodeable under this scheme. To verify my coding scheme, another member of our research group independently classified the questions under this scheme. His classification agreed with mine, with the exception of questions 1 and 2 (discussed above) for which he did not see an obvious classification.

Of the 12 questions that fell in the category of force questions there were no significant differences ($p < 0.01$) between the proportion of students answering correctly in the animated and traditional conditions. Four out of ten of the motion questions had significant differences in the proportion of students answering correctly and two of seven (30%) of the path questions had significant differences in the proportion of students answering correctly.

Interaction Required with Animation

In scheme two, the questions were classified by the level of interaction (with the animation) required to answer the question. The categories were: Question could be answered without viewing the animation (4, 9-11, 13, 15-17, 22, 24, 25, 28-30), animation had to be played (1-3, 6-8, 12, 14, 19-21, 23, 26, 27), and only the first frame of the animation had to be viewed (5, 18). The last category came about because these two animations required static information about the relative location of points of interest, but dynamical information was not needed. Two other members of our research group also classified the questions independently under this scheme. Their classification was identical to mine.

All six questions for which significant differences were found in the students' ability to answer correctly were from the group where the animation had to be played. The questions for which dynamical information was not needed to answer correctly were not affected by the animation.

Effect of Pictures

Finally, in scheme three, the questions were classified based on the pictures in the traditional version of the FCI. Some questions had accompanying graphics in the original version (5-12, 14-24, 28) and others had no original picture (1-4, 13, 25-27, 29, 30). Of the 20 questions with original pictures, 20% were significant. Likewise, of the 10 without pictures originally, 20% were significant.

Discussion

It seems that whether or not there was a picture in the traditional format is not as important as the topic of the question or the interaction required with the animation. The fact that students only responded differently when both dynamical quantities and the animation were relevant has an important implication.

Our society often views a technology rich classroom as being a better classroom. This can sometimes lead to a technocentric approach to education where classroom activities are centered on the technology rather than on the needs of the student. As a result, multimedia may be used to add flashy graphics or computer animation even when the addition is not central to the message. The animation sometimes improved performance, sometimes hindered performance, and often had no effect on performance. These results indicate that such an addition may be wasteful since superficial additions and changes in the problem statement did not change the students' responses.

No clear pattern emerged from this analysis regarding the kinds of question where the animation improved or hindered student performance. This issue will be taken up later in chapter 4.

3.4 Analysis of ACT Scores

The ACT is a standardized test that is "designed to assess high school students' general educational development and their ability to complete college-level work."[1] Many of the students in Sample 2 submitted their ACT scores to the university as part of the admission process. The university was able to provide English and Math ACT scores for these students. There were 241 students who answered at least 29 FCI questions and for whom ACT data was available. These students were used for the analysis in this section.

English and Math ACT Scores

The official ACT website[1] offers the following description of the English test.

> The English Test is a 75-question, 45-minute test that measures your understanding of the conventions of standard written English (punctuation, grammar and usage, and sentence structure) and of rhetorical skills

> (strategy, organization, and style). Spelling, vocabulary, and rote recall of rules of grammar are not tested. The test consists of five prose passages, each of which is accompanied by a sequence of multiple-choice test questions. Different passage types are included to provide a variety of rhetorical situations.

And their description of the math test is,

> The ACT Mathematics Test is a 60-question, 60-minute test designed to assess the mathematical skills students have typically acquired in courses taken up to the beginning of grade 12. The test presents multiple-choice questions that require you to use reasoning skills to solve practical problems in mathematics. Knowledge of basic formulas and computational skills are assumed as background for the problems, but complex formulas and extensive computation are not required.

The ACT English score is similar to the SAT verbal score and the ACT math score is similar to the SAT math score.

The Correlation Coefficient

A correlation coefficient is a measure of how much two measures are related. It ranges from -1 (those who were high on one measure were low on the other) through 0 (no linear relationship between the measures) through 1 (those who were high on one measure were also high on the other).

The Pearson product-moment correlation coefficient (*r*) is calculated by the following formula.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

Where

x, y = scores on each measure for a particular person

$\bar{x}, \bar{y}$ = the average score on each measure

Once r is calculated, significance can be tested by computing a z-score and then comparing the z-score to the normal distribution. For correlation coefficients the z-score is

$$z = (Z_r - Z_a)\sqrt{n - 3}$$

Where

Z_r = the Fisher's Z-transformed value corresponding to the sample value of r

Z_a = the Fisher's Z-transformed value corresponding to the hypothesized value of r

n = the sample size

The Fisher's Z-transformation can be looked up in a table in many basic statistics textbooks.[2]

The Analysis

Correlation coefficients were calculated between English/Math ACT scores and the FCI scores. The FCI scores were broken down into the separate components of force score, path score, and motion score as discussed in section 3.3.2. The results are presented below. Table 3.3 gives the correlation coefficients between English score and the FCI scores and Table 3.3 gives the same correlations with Math score. For this analysis, results are reported as significant if $p \geq 0.05$. A higher level of significance was used in the previous section because of the large number of statistical tests performed in that analysis. In this case only a few tests were performed so the 0.05 level can be used without a high number of expected false positives.

Table 3.2 - Correlations between English ACT score and FCI scores

English	Force	Path	Motion	Total FCI	N
Traditional	-0.012	0.138	0.347**	0.215**	158
Animated	-0.024	-0.056	0.181	0.066	83

* $p > 0.05$, ** $p < 0.01$ that $r \neq 0$ on a two tailed test

Table 3.3 - Correlations between Math ACT score and FCI scores

Math	Force	Path	Motion	Total FCI	N
Traditional	0.004	0.287**	0.380**	0.300**	158
Animated	0.188	0.187	0.371**	0.357**	83

* $p < 0.05$, ** $p < 0.01$ that $r \neq 0$ on a two tailed test

English ACT score was significantly correlated with the questions concerning motion and with the total FCI score for the traditional students. There were no significant correlations with English score for the animated group. These results indicate that part of what the traditional FCI measures is verbal ability while the Animated FCI probably does not. It is particularly striking that the correlation was lower for both the path and motion questions in the animated version. As discussed in section 3.3.2, these are the questions for which the greatest differences were seen between the two groups.

Apparently, performance on the animated questions is less linked to verbal ability than their traditional counterparts. We already have adequate measures of a student's verbal ability (such as the ACT) so we do not need to measure it again with the FCI. Since the animated questions are less linked to verbal ability, they have an advantage over the traditional questions in this matter. More will be discussed about the implications of this issue in the next chapters.

The animations do not seem to have an effect on the correlation between math score and FCI performance. Both the traditional and the animated FCI were correlated with math ability. These results seem reasonable. The animation questions require less reading and interpreting of words so it is not surprising, in hindsight, that animation performance was less correlated with verbal skills than the traditional questions.

However, there is no theoretical reason to believe that math ability would be less important in answering animated questions. The data seem to bear this out.

3.5 Analysis of Gender

There were 300 students who answered at least 29 questions, took either the fully animated or the traditional FCI, and had gender information available. The means and standard deviations, on several measures, from this group are shown in Table 3.4. The data is separated by gender and type of FCI taken.

Table 3.4 - Means and standard deviations for men and women on the two versions of the FCI and the English and Math sections of the ACT. ACT data was not available for all students. The data for ACT scores is based on 120 Trad Men, 69 Anim Men, 38 Trad Women, and 15 Anim Women.

		Men		Women	
		Trad	Anim	Trad	Anim
N		155	79	46	20
Force	Mean	3.02	3.06	2.17	1.85
	SD	2.20	2.56	2.12	1.60
Path	Mean	4.50	4.94	3.15	3.80
	SD	1.72	1.48	1.41	1.51
Motion	Mean	4.81	4.35	3.39	3.35
	SD	1.88	2.10	1.81	1.95
Total	Mean	13.0	13.1	9.11	9.70
	SD	4.47	4.77	3.64	3.93
English	Mean	23.7	24.5	25.8	26.9
	SD	4.09	4.34	3.96	3.60
Math	Mean	27.1	27.6	27.4	28.2
	SD	3.94	3.51	2.91	3.40

t-test for Population Means

The means were compared using a *t*-test assuming variances were unknown but equal. The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$\bar{x}_1, \bar{x}_2$ = the two means
n_1, n_2 = number of data points in the two populations
and s_p^2 is the pooled variance, given by

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

where s_1, s_2 = the two standard deviations
and degrees of freedom = $n_1 + n_2 - 2$.

Once the *t*-value is calculated, a *t*-table can be used to determine significance. In this case a two tailed-test should be used with the $p = 0.05$ level given by $t = 1.96$ and the $p = 0.01$ level given by $t = 2.58$. As with the analysis of ACT scores, the 0.05 level of significance is reported because the number of statistical tests performed in this analysis is low enough that a large number of false positives would not be expected.

Comparison of Men and Women

t-tests were performed to compare the means of men and women on the various measures on both versions of the FCI. Table 3.5 lists the *t*-values for these comparisons.

Table 3.5 - *t*-values found when the means of men and women on various sections of the FCI were compared. Values for English and Math ACT were calculated with the animation and traditional group combined. There were no significant differences between the animation and the traditional group on the ACT measures.

	Trad	Anim
Force	2.31*	2.02*
Path	4.85**	3.05**
Motion	4.52**	1.94 (p=0.06)
Total	5.41**	2.97**
English	-3.31**	
Math	-0.59 (Not Significant)	

*$p < 0.05$, **$p < 0.01$

These results show that men outperformed women on all FCI measures, but women had higher English ACT scores.

Comparison within the Genders

t-tests were also performed to compare the means on the various measures between men on the two versions and between women on the two versions. The results are given in Table 3.6 below.

Table 3.6 - *t*-values found when the means of each gender on the traditional and animated versions were compared. A negative value indicates that the mean of the animated group was higher. NS = not significant.

	Comparison of Men	Comparison of Women
Force	-0.14 (NS)	0.61 (NS)
Path	-1.91 (p = 0.06)	-1.68 (NS)
Motion	1.67 (NS)	0.08 (NS)
Total	-0.20 (NS)	-0.59 (NS)

None of the measures produced statistically significant differences if 0.05 is taken as the level for significance. However, the difference between path score for men on the two versions had a *p*-value of 0.06.

Correlation with Gender

Correlation coefficients were also calculated between gender and ACT scores, as well as between gender and FCI measures. Males were given the value 0 and females were given the value 1 for this analysis. So a negative correlation indicates that men generally had higher scores. The results are shown below in Table 3.7.

Table 3.7 - Correlation with gender

correlations	gender	
	Trad	Anim
English	0.22*	0.21*
Math	0.03	0.07
Force	-0.18*	-0.19*
Path	-0.38*	-0.24*
Motion	-0.32*	-0.08
Total	-0.39*	-0.22*

*Significantly different from zero ($p < 0.05$ on a two-tail test)

There are several important things to note in these results. Gender is an important factor on all of these measures except the ACT Math score. Women generally did better than men on their English ACT and men generally did better on all aspects of the FCI. The one exception was the motion questions on the animated FCI. Two correlation coefficients can be compared using a *Z*-test given by

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

where n_1, n_2 = the two population sizes
and Z_1, Z_2 = the two Fisher's *Z*-transformed values corresponding to the values or r_1 and r_2.

When the correlation coefficients between gender and the motion score on the two versions of the FCI are compared in this manner, $Z = -1.99$ which is significant on a two-tailed test at the $p = 0.05$ level, which is the level commonly used for tests of significance. In other words, motion is significantly more correlated with gender on the traditional FCI than on the animated FCI.

Discussion

Although the analysis presented in this section is straightforward, gaining insight from the results is a bit more difficult. One thing that is very clear is that men generally perform better than women on all aspects of the FCI and on both versions. This is a disturbing result, especially in light of women's superior performance on the ACT English test and their comparable performance on the ACT Math test.

Although others have noticed the discrepancy between the genders[3], the reasons for the difference are not understood at his time. In a study that is currently unpublished, I investigated the relationship between performance on the traditional FCI and previous involvement in sports. I found that sports playing was positively correlated ($r = 0.30$, $p < 0.06$) with pre-test FCI performance for university women taking calculus-based mechanics. When only those women whose scores fell below the suggested "threshold of understanding" of 60% were considered, the correlation jumped to $r = 0.56$ ($p = 0$). This indicates that the relationship between sports playing and FCI performance for women drops as women's knowledge grows. I found no correlations with sports playing and FCI performance for the men. My results are similar to Hanson[4], who found a positive correlation between science performance and involvement in sports for white girls but not for boys or African-American girls.

Although this link between sports playing and FCI performance does provide a direction for future research, it does not explain the difference seen in the genders. Unfortunately, it is beyond the scope of this project to explain why women as a whole do not do as well on the FCI as men. However, one important question that must be asked from this study is "Does the animation affect the genders differently?"

At first glance, there does not appear to be much of a difference for men and women. After all, men still outperformed women on all but one aspect of the animated FCI, at a level that was statistically significant, and statistically significant results were not found within the genders except on the path questions for men. And finally, the correlations between gender and performance on the various groupings of the FCI were all significant except for the motion questions on the animated FCI.

However, it can not be concluded that there is no interaction between gender and animation. Although women performed significantly below men on the animated as well as the traditional FCI, the gap was consistently smaller on the animated FCI. In other words, the difference between the performance of men and women on the traditional FCI is greater than the difference in performance on the animated FCI. What is particularly striking is that the gap was smaller on all the measures; force, path and motion.

It is also interesting to note that the correlation between gender and performance on motion questions was significantly higher for the traditional FCI than for the animated FCI. In fact, the correlation between gender and performance on motion questions dropped to zero in the animated condition. As was discussed in Section 3.3.2, the motion questions are closely tied to the animation. The force questions are not closely tied to the animations and it was on those questions that the least differences were seen regarding gender.

It is difficult to draw out implications from these results. The animations probably decrease the difference in performance between men and women though the effect is not dramatic. However, this in and of itself is not necessarily positive. A careful review of the motion scores shows that the gap decreased because women performed about the same in the traditional and animated conditions but men actually performed worse on the animated questions. So the gap decreased because men lost ground. Until it is better understood why there is a difference between men and women on FCI performance it will be difficult to give much meaning to the results presented here.

3.6 Test and Test Item Analysis

There are several methods available for analyzing a test, after it has been constructed, which give important insights into the quality of the test. In this section, I will discuss the results of reliability, difficulty, and discrimination analysis for both the traditional and animated versions of the FCI.

3.6.1 Reliability

Ideally, a test should yield the same scores every time it is given to an identical population. In reality, there is always some measurement error in every test. The reliability coefficient, which ranges from 0.0 to 1.0, indicates how consistant the test is at measuring whatever it measures. A reliability coefficient of 0.0 would indicate that the test does not consistently measure anything and a coefficient of 1.0 would be a perfect test, with no measurement error. Doran[5] offers the following guidelines for interpreting reliability coefficients for educational testing purposes

0.95-0.99	Very High, Rarely Found
0.90-0.95	High, Sufficient for Measurement of Individuals
0.80-0.90	Fairly High, Possible for Measurement of Individuals
0.70-0.80	Okay, Sufficient for Group Measurement, Not individuals
Below 0.70	Low, Useful Only for Group Averages or Surveys

There are a number of methods available for calculating the reliability coefficient. One of the most popular is the Kuder-Richardson 20 which can be found in many measurement and evaluation textbooks.[7] This statistic is based on the mean correlation between all items on the test. In other words, it checks the internal consistency of the test. The formula for the KR-20 is

$$\mathrm{KR}(20) = \frac{K}{K-1}\left(1 - \frac{\Sigma[p(1-p)]}{\sigma^2}\right)$$

Where

K = the number of test items

p = the proportion passing a given item

σ = the standard deviation of the total scores

The KR-20 reliability coefficient was calculated for both the animated and traditional versions of the FCI based on the data collected for this study. Since the students in Sample 1 took a mixed version of the FCI their scores were not used for this analysis since reliability analysis looks at a test in its entirety. Also, question 10 was eliminated from the analysis due to problems with data collection as already discussed. As a final note, only those students who answered 29 or 30 of the questions are included in the analysis. Those students who did not finish the test were removed. The results are shown in Table 3.8 below.

Table 3.8 - KR-20 reliability coefficients for the two versions of the test. *N* is the number of students contributing to the analysis.

	KR-20	*N*
Animated FCI	0.79	99
Traditional FCI	0.75	215

These results indicate that the AFCI is slightly more reliable than the traditional FCI. However, the difference is small and may be the result of sampling error. Neither test is reliable enough to be used for individual assessment but is acceptable for group measurement. Since the FCI is generally given to a class of students, either version would be acceptable.

Curiously, the reliability coefficients found here differ from what has been reported elsewhere. As was discussed in Chapter 2, the authors of the FCI claim a reliability coefficient of 0.86 for the pre-test and 0.89 for the post-test on the Mechanics Diagnostic Test, which they claim is similar enough to the FCI that recalculation of reliability coefficients is not necessary. Others[6] who have calculated a KR-20 for the latest version of the FCI have found values around 0.84-0.86 on both the pre and post-tests (traditional version).

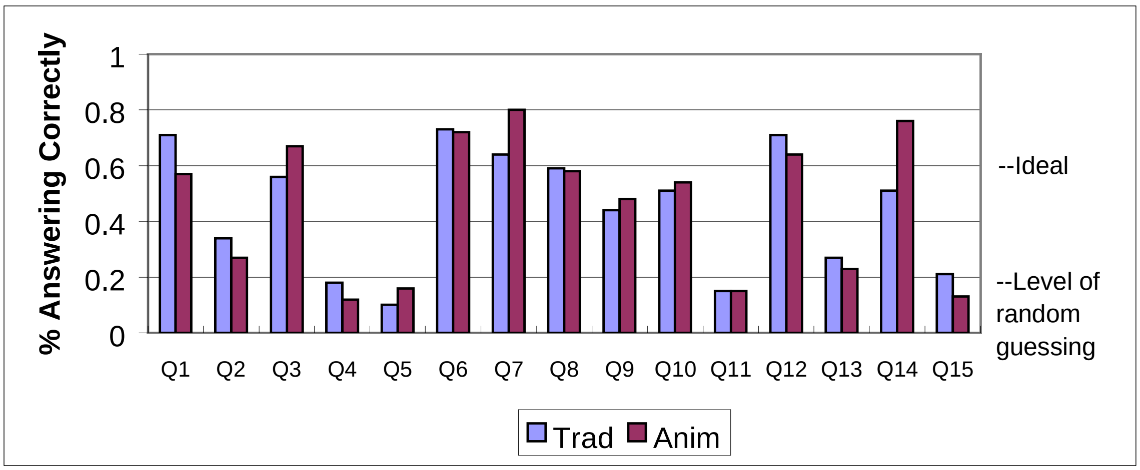
It is not clear why higher reliability coefficients have been found in other places and it is beyond the scope of this study to fully investigate the issue since this question can not be answered with the data collected as part of this study. The most likely explanation is that reliability of the test is affected by the importance placed on the test by the proctors. The locations reporting high reliability tend to place more value on the results than is typical at the location of my data. In any event, the reliability coefficient found in this study is sufficiently high to be used for group assessment, and the difference between the animated and traditional versions was small.

3.6.2 Item Analysis - Difficulty

The difficulty of an item is defined as *R/N* where *R* is the number of students answering correctly and *N* is the total number of test takers. In other words, difficulty is the percent of students answering a question correctly. Ideally, each item should have a difficulty mid-way between 100% and random guessing. A difficulty at this level leaves the most room to discriminate between students. If the difficulty is too high or too low then many students will be clumped together at either the high or low extreme and it will be impossible to distinguish between these students.

For a test with 5 answer choices for each question, random guessing would give a score around 20% so the ideal difficulty would be around 0.60. Questions which are very easy or very difficult should be avoided in test construction. These questions do not give much information if a comparison of students is desired.

An analysis of difficulty was carried out using all available data from both Sample 1 and Sample 2. The results are shown in Figure 3.2 below.



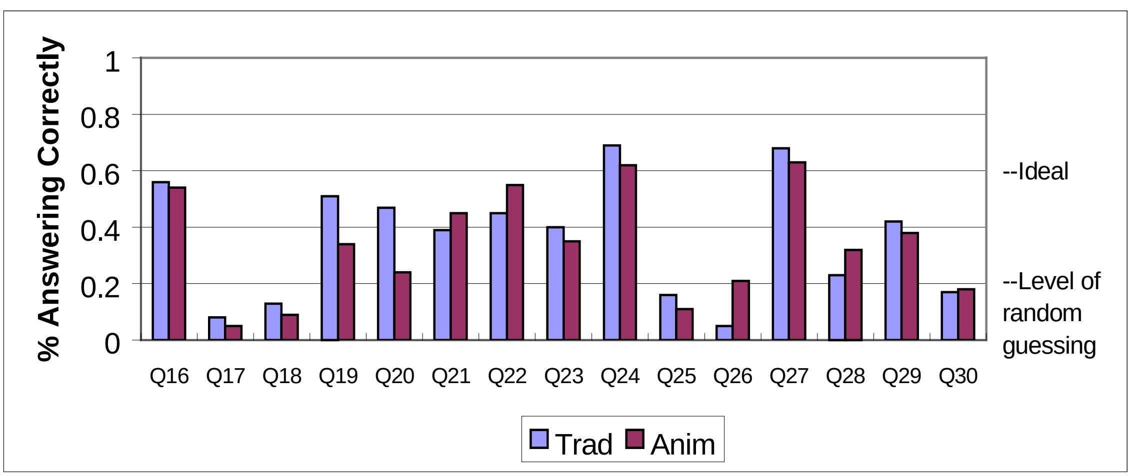


Figure 3.2 - Item difficulty (% Correct) for each question. Shown for both the traditional and animated versions of the FCI.

Eleven traditional questions and 12 animated questions had a difficulty index less than 0.30 which is in the range of random guessing. This does not necessarily mean that students randomly guess at these questions, just that the scores are indistinguishable from random guessing.

Normally, a test with over one-third of the questions in this category would be a cause for concern. However, in this case it is not since the data is based on pre-test scores. Although these items may not be useful in a pre-test situation, they may still be valuable items over all. Since the FCI is also used as a post-test (and these are the scores that are usually of most interest) low difficulty scores on the pre-test are not necessarily a problem. There are also indications that students are not randomly guessing but have strong misconceptions. This has already been discussed in Chapter 2 and will be discussed further in the next chapter.

Questions with a difficulty index in the range of 0.3 to 0.7 are probably ideal for a pre-test that will also be used as a post-test. Sixteen traditional questions and 15 animated questions fell into this category. This numbers are comparable, and the fact that they account for about half of each test is encouraging.

There were three traditional questions (#1, #6, and #12) that more than 70% of the students answered correctly. This is an indication that these questions are not large discriminators on the pre-test and would likely be even worse on the post-test. Likewise there were three animated questions with the same concern (#6, #7, and #14). More will be discussed about this issue in the next section and in the discussion of the qualitative data in Chapter 4.

3.6.3 Item Analysis - Discrimination

The discrimination index is a measure of how well an item separates those students who did well on the test as a whole and those students who did poorly on the test as a whole.

Items with a large discrimination index are generally better questions because they do a better job of distinguishing students. The discrimination index can range from -1.0 to 1.0. Negative values are an extreme cause for concern because they indicate that the poor students answered the question correctly while the good students answered incorrectly. Negative values often indicate that a question has been graded incorrectly. A discrimination index of 1.0 indicates that all the students in the high group, in overall score, got the question correct and all the students in the low group missed the question.

There are various methods of choosing the high and low group; none of which stand out as being better. I chose to form my groups in the following manner. The students were ranked by their total score and divided into thirds based on performance on the whole test. Those students who had the same score as students in the low or high group were moved from the middle group to the appropriate group. For example, after the groups were divided evenly the low group from the animation condition contained 33 students with scores ranging from 3 to 10. However there were 2 students who fell in the middle group who also had total FCI scores of 10, so those two students were moved to the low group.

After this was done the low group was comprised of students with a total FCI score of 0-10, the middle group had students with scores of 11-13, and the high group had students with scores 14-30. Only those students who had a total FCI score based on having answered at least 29 questions were used in this analysis. The total numbers of students making up each group is shown in Table 3.9 below.

Table 3.9 - Number of students making up the low, middle, and high groups.

N's	**Low**	**Middle**	**High**
Traditional	93	44	78
Animation	35	29	35

Once the groups are formed, the discrimination index is computed according to the following formula.

$$D = \frac{H}{N_H} - \frac{L}{N_L}$$

Where,

D = discrimination

H = number of students in the high group choosing correct answer

L = number of students in the low group choosing the correct answer

$N_{H,L}$ = total number of students in high, low group

The results are reported in Figure 3.3 below.

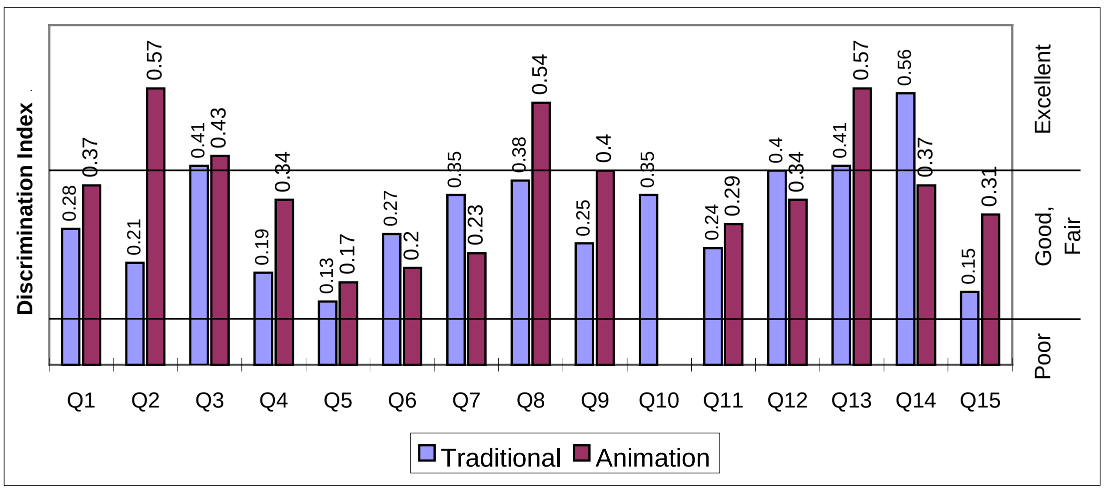


Figure 3.3 - Item discrimination for the traditional and animation group. The data is taken from Sample 2 only. As previously discussed, data is not available for the animation group on Question 10.

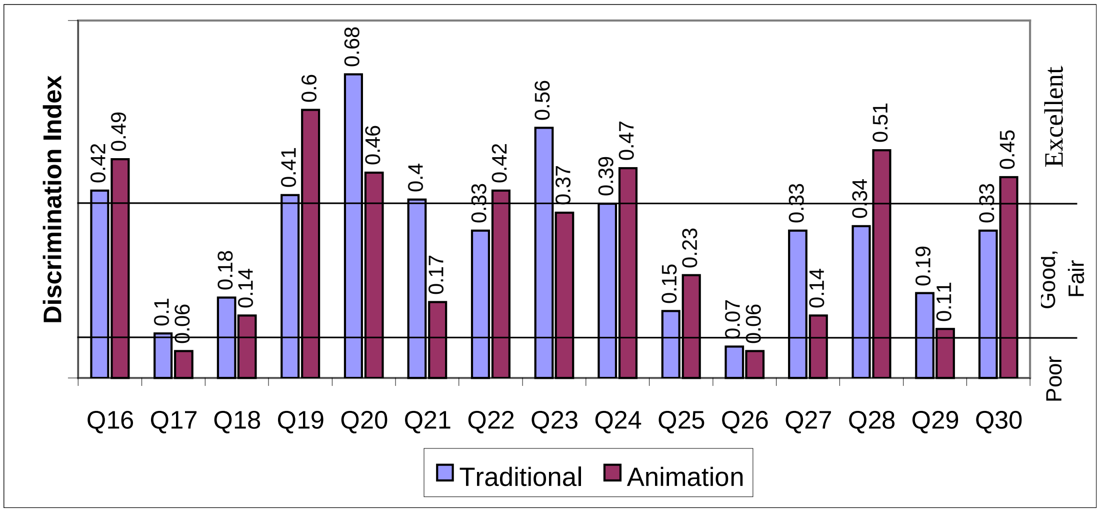


Figure 3.3 - Continued

No clear overall patterns emerge from this data. The average discrimination was essentially the same in both groups. The traditional group's average discrimination was 0.32 and in the animation group it was 0.34. Based on the following criteria,[7] both of the numbers are good.

Index of Discrimination	Evaluation
0.40+	Excellent discrimination
0.30-0.39	Good Discrimination
0.10-0.29	Fair Discrimination
0-0.09	Poor Discrimination
Negative	Item may be miskeyed or intrinsically ambiguous

About one-third of the test-items from both the traditional and animated tests have excellent discrimination. Very few fall in the category of poor discrimination. However, there are some items with very low discrimination. There are also several items for which the difference in discrimination is large. These items are very interesting and will be discussed in more detail in the next chapter.

3.7 Chapter Summary

In this chapter I have presented the quantitative data collected as a part of this project. To summarize, the main findings are,

- Animation can affect the way a student will answer a conceptual question like those found on the FCI.
- An effect is most likely when the question involves dynamical events and when the animation is central to answering the question.
- Animation can either improve or hinder performance.
- Performance on animated questions is less linked to overall verbal ability than performance on traditional questions.
- Animation does not change the relationship between performance and math ability.
- Animation may reduce the gap in scores between men and women.
- Both the animated and the traditional FCI are sufficiently reliable.
- Analysis of difficulty and discrimination indicates that some questions (on both versions) should be looked at more carefully. This will be done in the next chapter.

[1] http://www.act.org

[2] Kanji, G. K. (1993). 100 Statistical Tests. London: Sage Publications.

[3] McCullough, L. (1999). Personal Communication.

[4] Hanson, S. L., & Kraus, R. S. (1998). Women, sports, and science: Do female athletes have an advantage? Sociology of Education, 71, 93-110.

[5] Doran, R. L. (1980). Basic Measurement and Evaluation of Science Instruction. Washington, DC: National Science Teachers Association.

[6] Henderson, C. (2000). Personal Communication.

[7] Hopkins, K. (1998). Educational and Psychological Measurement and Evaluation (Eighth Edition). Boston: Allyn and Bacon .

Chapter 4
Analysis of the Items

In the previous chapter, I presented and discussed the quantitative data collected. Although this data provides many interesting insights and shows that students can be affected by animation, the data fails to adequately explain *why* the animation had an effect. Without answering the question of "Why?" it is difficult to make recommendations regarding conceptual assessment. In this chapter, I will attempt to offer explanations for the choices students made.

4.1 Qualitative Research

Traditionally, educational research has focused on quantitative research. Patterned after the highly successful research methods of the physical sciences, these methods follow well-defined rules for data collection and analysis. When done well, quantitative analysis can produce results that are easily replicable and generalizable with precisely known limitations. However, quantitative methods often obtain generalizability at the expense of detail.

Qualitative methods are excellent for collecting lots of detailed information. For practical reasons, this detail is usually found by focusing on a small number of informants. In other words, instead of retrieving a little bit of information from many people, a lot of information is obtained from a small number of people. The trade-off, of course, is generalizability and the rigor associated with statistical analysis.

I believe that the best educational research uses both qualitative and quantitative methods. The quantitative results can be used to generalize and find replicable trends, and then qualitative analysis can be used to explain the quantitative results and to guide future research. I have used this philosophy as a basis for this research project.

4.2 Think-Aloud Interviews

I wanted to know what students were thinking as they answered the FCI questions. It is not possible to know why students chose a particular answer just by looking at the answer they gave. I also found that students did not articulate, in enough detail, the reasons behind their choices when asked to write a brief reason. Patton[1] offers the following description of interviewing.

> The purpose of interviewing is to find out what is in and on someone else's mind. The purpose of open-ended interviewing is not to put things in someone's mind (for example, the interviewer's preconceived categories for organizing the world) but to access the perspective of the person being interviewed. We interview people to find out from them those things we can not directly observe.

Because I wanted to know about things in the students' minds that could not be directly observed, I conducted interviews with 15 students during the Spring of 2000. All 15 students were enrolled in first semester calculus-based physics at a large university in the South. The students were each paid $10 for participating. At the beginning of the semester I sent an e-mail requesting volunteers to all of the students of one lecturer. I later recruited individual students by calling them on the telephone.

In general, I avoided calling students who had received an "A" or a "F" on their first exam, as I believed average students would yield more useful information. I did not want very high ability students because wrong answers are generally more interesting than right answers. Also, I already had a feeling that the students who had volunteered at random represented an above average group. I avoided the very low performing students for several reasons. One, I didn't want students who were not participating in the course. Two, they are an extreme group and I was limited in the number of students I could interview. And finally, I had already interviewed one very low scoring student who volunteered at random.

One student was removed from the analysis because he admitted he had taken the FCI twice while in high school. So my actual sample of analyzable interviews was 14.

4.2.1 Format of the Interviews

Each interview was scheduled at the convenience of the student and took place in a room on campus. The interviews were conducted in private with only the student and myself present. The students were asked to sit at a table with a laptop in front of them. I sat next to the student being interviewed.

Before the interview began, the procedure for the interview was explained to the student and he/she was asked to sign an informed consent. (See Appendix C) A video recorder was placed just behind the student so that interactions with the computer were captured on video and statements made by the student were captured in audio.

Early Format

For the first three interviews, I asked the students to work through either Form A or Form B of the mixed version FCI (see Section 3.2) on their own. The students were then asked to write a brief reason for each answer they gave. Then, after the students had completed the test I asked them to discuss some of the items with me. The information I got from this style of interview was not very rich so I abandoned it.

Main Format

The remaining 11 interviews were conducted in the following manner. The students were asked to work through either Form A or Form B of the mixed FCI while talking out-loud. Specifically the students were instructed to say everything that they were thinking while answering the questions. I told the students that I was not testing them, that I was interested in what they thought as they answered the questions and in why they made the choices they made. In general, I was quiet while the students worked. If they were silent for a long period of time, I would ask them what they were thinking about in an effort to get them talking again. If it was not clear to me why a student chose a particular answer, I would ask the student to explain his/her reasoning again.

In seven cases, I asked the student to look at the alternate form after the entire test was completed. In other words, if a student took Form A, he/she was asked to look at Form B upon finishing Form A. In these cases I asked the students to look at particular questions and asked them if they wanted to change their answer. I did not usually have them relook at all the questions because of time constraints and student fatigue. The questions I asked them to look at again were ones in which I felt, based on their comments and my intuition, that they may answer differently in the opposite form (animation or traditional). The choice of questions was different for each student. I made a point of including at least one question they had originally answered correctly. I specifically told them the review questions were questions for which I had a particular interest and not necessarily ones in which they had answered incorrectly. I asked the students to go back and look at particular questions in the opposite form because I thought their responses would be insightful. And, it turned out I was right.

End of the Interview

After students were finished working through the FCI they were asked to fill out a background questionnaire (See Appendix D). The questionnaire was used to determine general characteristics of the interviewees.

Interviewer Bias

It is impossible to remain detached in qualitative research. My own experiences, philosophies, and personality have certainly shaped the outcome of this study. I have an expert understanding of the concepts tested on the FCI. I do not share the misconceptions that are found among students and in many cases, I do not ever remember having these misconceptions myself. Therefore, I have had to pay particular attention to understanding the students' perspectives of our physical world.

I also have experience teaching in a physics classroom and am very familiar with the common student misconceptions documented in the physics education research literature. Because of this experience, there were responses that I was expecting to see as students worked their way through the questions. I have tried to look beyond my expectations as I analyzed the interview transcripts.

4.2.2 The Students

Below is a summary of the information I was able to collect about each student. The FCI score is based on the think-aloud interview and the Test 1 score was obtained from the professor of the course. All other data is self-reported.

Total number of interviews analyzed = 14
Number of Males = 7
Number of Females = 7
Number who had never taken physics before = 1
Average Math SAT = 664 (range = 530-800)
Average Verbal SAT = 628 (range = 460-760)
Average FCI score = 14 (range = 3-27)
Average Test 1 Score = 75 (range = 37-100)

The students represent a reasonable range of what would be expected in the whole class slightly skewed toward higher ability students. There were also more than expected female interviewees.

4.2.3 Analysis of the Interviews

Fourteen interviews were entirely transcribed. (As noted earlier, one interview was discarded due to the student having taken the FCI twice before). The transcriptions included notations to indicate how and when the students interacted with the computer. Appendix E contains one transcribed interview as an example.

Because I was most interested in the questions, and not individual students, a cross-case analysis was performed. In other words, the interviews were analyzed by question instead of by student. For each question, students' responses regarding that question were read and notes were taken. I grouped students together based on whether they saw the animated or traditional form and by the correctness of their initial answer. For each student, I recorded my own summary of their thinking on that question and made notes of any unusual or striking aspects of their response.

My analysis at this point was inductive. Patton[1] states that "Inductive analysis means that the patterns, themes, and categories of analysis come from the data; they emerge out of the data rather than being imposed on them prior to data collection and analysis." I followed this description as I went through my notes. Except for a focus on the effect of the animation, I did not have specific categories in mind as I looked through the data. I simply read carefully, thought hard, and tried to be open minded enough to allow relevant points to emerge from the interview data without restrictions.

This type of analysis is also called grounded theory. Strauss and Corbin[2] offer the following definition of grounded theory.

> A grounded theory is one that is inductively derived from the study of the phenomenon it represents. That is, it is discovered, developed, and provisionally verified through systematic data collection and analysis of data pertaining to that phenomenon…One does not begin with a theory, then prove it. Rather, one begins with an area of study and what is relevant to that area is allowed to emerge.

I chose grounded theory as an analysis method because I did not want to limit myself in outcome. I wanted to allow ideas that I could not have envisioned to emerge.

The "relevant points" I discuss were developed from the interview data without looking back at the results of the quantitative analysis. After I believed I had an understanding of various aspects of each question I then checked for consistency with the quantitative results. If the results were consistent then it is noted in the discussion. Likewise, if the quantitative results were not consistent then either the theory was revised or the discrepancy is discussed.

Presentation of Interview Findings

I have tried to present my interview findings in a way that will allow each reader to made his or her own judgements. For example, I have included as many direct student quotes as was reasonable and I have always backed up important assertions with such quotes. Unfortunately, it is impossible to include all of the students' statements due to the large volume of data I collected.

Quotes are sometimes edited for clarity since it is not practical to include every long statement made. When words are removed they are replaced with "…". Sometimes, I have added words to make the meaning clearer, they are contained in square brackets. So for example, the statement "I'm not sure if it applies here, but that is gonna, it's gonna accelerate and then stay at a constant velocity while it's falling." Could be rewritten as "I'm not sure if [terminal velocity] applies here … it's gonna accelerate and then stay at constant velocity while it's falling". I have been careful to not change the meaning of any of the statements but to simply present them in a clear and concise way.

4.3 Discussion of the Test Items

The result of analysis of each question is presented in the sections that follow. A graph showing the distribution of responses is shown followed by a summary of statistics for that particular question. Note that the data presented in the graph and table are from the large scale data collection. They do not include responses from the interviews. Details of the statistical data can be found in Chapter 3.

After a summary of the statistical data, there is a discussion of the findings from analysis of the interviews. When students are quoted, they are given names to identify them as individuals. The names given for the interviewees are not their real names but they do accurately reflect the gender of the student.

Discussion of the interview data is followed by a discussion of the effect of the animation which is based on both the statistical and interview data. And finally, I make a recommendation for use of that particular item for conceptual assessment. It should be noted that my data is almost entirely based on a university population taking calculus-based mechanics and answering the FCI as a pre-test. The recommendations I give therefore apply specifically to this type of student. However, I believe my recommendations are useful beyond this one application.

As a final note, remember that copies of both the animated and the traditional FCI are found in Appendix A and Appendix B. The reader is urged to reference these appendices while reading the rest of this chapter as I assume familiarity with each.

4.3.1 Item One

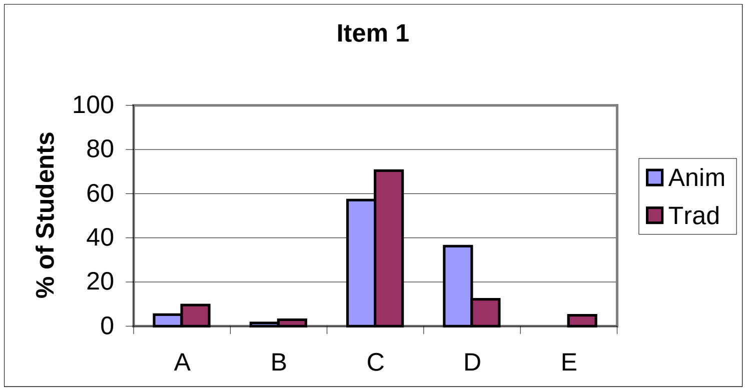


Figure 4.1 - Distribution of student answers for Item 1

Correct response = C

Choices with significant differences between animation and traditional = C, D, E

Table 4.1 - Summary of Statistical Data for Item 1

Item One	Traditional	Animation
Correlation with English ACT	0.15	0.06
Correlation with Math ACT	0.15	0.11
% Correct	71	57
Index of discrimination	0.28	0.37

Points of Interest from the Interview Data

- Students who answer correctly "remember" the answer from class.

What was most striking about this question was that the students who answered correctly did not seem to understand why the balls would fall at the same rate. Every student who answered correctly either specifically stated that it was something he/she remembered or gave an explanation that was incorrect or vague. Even the one student who had never taken physics before "remembered" the correct answer. When I asked her about it she told me it had been discussed in a general science class she had taken much earlier.

For example, after reading the question but before viewing the animations Tanner stated "I think that if air resistance is negligible, everything falls at the same rate, so they should hit the ground at the same time." When I later asked him how he knew that, he replied "Well, I learned that in class, I guess". Nathan, who saw the traditional form of the question answered, "It'll be about the same for both balls. That's one thing they grill into you." Three students specifically mentioned a bowling ball compared to a feather or a pebble and two others talked about Galileo.

- Students can answer correctly but have grossly incorrect ideas.

Interestingly, several students answered correctly but made statements that indicated that their understanding was poor. For example, consider Charlie who saw the animated form of the question. He stated that "falling is a representation of mass and air resistance" and justified the correct choice C with "because falling down is a force of air resistance, it's because of air resistance. Mass has nothing to do with it and since both the balls are the same size and the same shape, they would fall at the same rate." Presumably, these students were able to answer correctly, in spite of a weak understanding because they could rely on their memories of formal instruction.

- The conflict between the remembered response and true beliefs can cause turmoil for students.

Barb (saw animation) answered correctly but then stated "I've always been told that a feather and a bowling ball will fall at the same, you know I still have this little, you know it never settles very right but, I know that's what I've been told but at the same time, I'm like, I really want the heavier ball to fall faster." Although she knew the correct answer and responded accordingly, it is obvious that her response is not a measure of her true beliefs. This same issue was seen in Mazur's classroom upon giving the FCI and is discussed in Section 2.1.5.

For Lucy, who saw the traditional version, it was not that simple. Although her thoughts are very verbose, they are insightful and are partly quoted below. I have added bold to draw attention to particular aspects.

> Lucy: If you throw something and you drop something, they end, they hit the ground at the exact time, cause **I remember that from that physics, my physics class**, this demonstration. But, I can't remember if that depends on, one weighs, I mean obviously if you drop a feather and a bowling ball, the feather's gonna hit the ground later, but that's because of air resistance. Is that because of the weight, too? Or is that just because of air resistance? I'm gonna say, I think that weight and mass are somewhat different in this, so **I'm gonna probably go with about the same**. I think I would.
>
> M: Ok. And why is that?
>
> Lucy: Because I can't**, I know that if they have the same mass that they would hit the ground at the same**. The obvious assumption would be that the one with the greater weight would hit the, actually no. **I don't think it is about the same for both**. Two metal balls are the same size, one weighs twice. Ok, if one weighs as much as the other, that would be like saying one has twice as much mass or, yeah, twice as much mass, so no, it wouldn't be this one, it would be, the one with twice as much mass**, "the time it takes the balls will be half as long for the heavier ball as for the lighter one"**. I'd say that. Because the heavier one is gonna hit the ground, since it weighs twice as much it's gonna have**, they're both going to be pulled at the same rate. 9.8** - Ok, but that's what I'm gonna say.
>
> M: Ok, and why is that?
>
> Lucy: Because the heavier ball is going to, which weighs twice as much as the other one, you would think that it would be pulled, **well they're both accelerating at the same rate,** I can't remember the equation though. There's an equation that would help me, like you know, for sure. But I think it's because they're both, mmm. I don't really know, it just makes sense. It kind of seems right. **It seems like if you threw, yeah a bowling ball versus something half the size of a bowling ball that the bowling ball would hit the ground first. It seems like that's right. It doesn't look right and I think the equation would tell me otherwise. But that's what I would think if I didn't have the equation with me.**

Lucy changes her answer several times during the discussion. She begins by relying on what she remembers from class but then changes her answer when thinking more about it. On a number of occasions she makes correct statements, such as "they're both going to be pulled at the same rate, -9.8", but does not let these ideas take her back to the correct answer. In the end she indicates that she is probably incorrect but her answer seems right.

- Some students believe heavier objects will fall faster.

Five of the students answered incorrectly and stated that the heavier ball would fall faster. They often discarded choice A in favor of choice D feeling that the heavier ball would not fall twice as fast, but only slightly faster. However, no student was able to offer an explanation for their choice. In general, students who answered Item 1 incorrectly based their answers on instinct.

Did the animation have an effect?

The quantitative data suggests that the animation affected the answers students gave. However, I did not see any clear indications of this in the interviews. If the effect of the animation is genuine, and not a statistical anomaly, then it is probably because the animation is less likely to bring out the memorized response that seems to account for the majority of correct answers. Perhaps students are less likely to answer by their memories and more by their beliefs when they see the question posed in a more intuitive way.

The quantitative data does back up this theory. If students were more likely to "remember" the correct response when answering the traditional question then,

1. It would be expected that more students would get the traditional question correct.
2. It would also be expected that the animated question would be less correlated with verbal skills since those with good verbal skills are probably the ones best able to use their verbal memories.

3. And finally, it would be expected that the animated question would be a better discriminator since those who did not understand would be less likely to answer correctly.

All three of these expectations were found to be true which gives strong support to the idea that the traditional version is more likely to bring out memorized responses.

Although the interview data does not clearly explain why a statistically significance difference was seen in the distribution of responses in the large scale data collection, it does indicate that the animated version of the question is probably a better test item. Students who answer correctly generally do so for the wrong reasons and the animated version does appear to reduce this tendency.

Recommendation - Discard Entirely

Based on all the data collected, I do not believe Item 1, in any form, is valuable for assessment purposes. A correct answer does not indicate much more than that the student has been exposed to, and can recall the information. If Item 1 must be used then the animated form would be preferred because it is a better discriminator and perhaps less likely to give false positives.

4.3.2 Item Two

Figure 4.2 - Distribution of student answers for Item 2

Correct response = A

Choices with significant differences between animation and traditional = B, D

Table 4.2 - Summary of Statistical Data for Item 2

Item Two	Traditional	Animation
Correlation with English ACT	0.19	0.08
Correlation with Math ACT	0.21	0.22
% Correct	34	27
Index of discrimination	0.21	0.57

Points of Interest from the Interview Data

- Students make the same choices in Item 2, for the same reasons, as in Item 1. However, there are fewer false positives.

To an expert, Item 2 is the same as Item 1. However, this similarity is not necessarily evident to the student as demonstrated by the rather large difference in correct answers given in the two questions. The key to answering both questions correctly is in understanding that gravity accelerates all objects the same, regardless of their masses.

Item 1 contains the same surface features as the standard feather and rock demonstration so commonly done and discussed in general physics courses. Although the concepts are the same in Item 2, it looks different. The balls fall in two dimensions and the question asks about distance covered, not about time of fall. Because Item 2 does match the superficial features of the standard demonstration, it would be expected that students would be more likely to answer according to what they believe and less by what they remember from class. The poorer scores on Item 2, compared to Item 1, support this notion.

The interviewees who answered Item 2 correctly all answered Item 1 correctly as well. They gave basically the same explanation for Item 2 as they did for Item 1. The students who answered both Item 1 and Item 2 incorrectly also tended to give the same reasons for both questions.

Interestingly, none of the three students who answered Item 1 correctly but Item 2 incorrectly gave the standard "heavier objects fall faster" as the reason for their choice in Item 2. Maggie stated that,

> "If she had a train and a basketball and they're both going the same speed, and like, one would, like the train is gonna take longer to stop…I think E because they're both falling about the same speed and, but the red one's going farther." (Maggie, Animation)

Maggie understood that they would both fall at the same rate vertically, but demonstrated a misconception about the horizontal motion.

Justin stated that,

> "Since force equals mass times acceleration, and I guess the same speed, if they have the same speed they have constant acceleration or something like that, so that being equal and the mass of the one ball being more, it seemed like it would have the higher horizontal force (so the one with more mass) would go further." (Justin, Animation)

Justin confused the effect of force and acceleration. He had correctly equated same acceleration with same time of fall in Item 1.

Barb's reason for her choice in Item 2 was "I just went and saw *Bringing Out the Dead*". She went onto explain how two men in the movie, one light and one heavy, had jumped from a window and landed in different spots. It should be noted that Barb was the student who answered correctly for Item 1 but specifically stated that she did not feel comfortable with her answer and only gave it because she remembered from class that it was correct.

Did the animation have an effect?

Besides the fact that fewer students answered Item 2 correctly, I did not see major differences in the way they responded to the two questions. It is not clear that they give the correct response for the correct reason, and I saw no clear difference between the animated group and the traditional group.

Recommendation - Animation

Item 2 does eliminate many of the false positives seen in Item 1. This is probably because it is superficially different from the standard demonstration and discussion. Just as with Item 1, the animated question is less correlated with verbal skills, a better discriminator, and more difficult. And again, it is likely that the animated Item 2 is less likely to produce false positives due to memorized responses than the traditional version of Item 2. I believe that Item 2 is a fair measure of students' ideas and that the animated version is superior for assessment purposes. This is strongly supported by the excellent discrimination found for the animated form and only a fair discrimination found for the traditional form.

4.3.3 Item Three

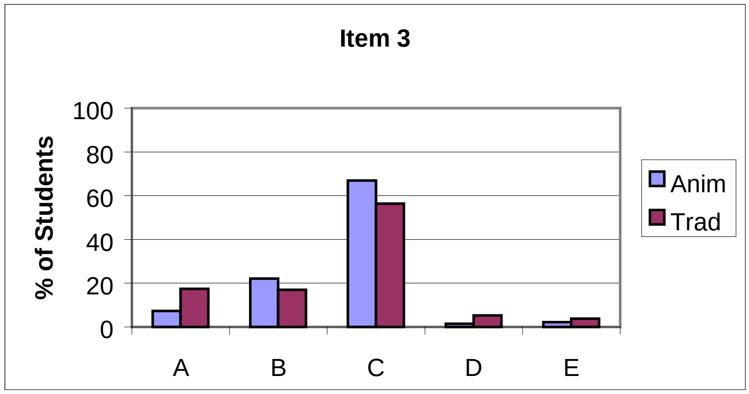


Figure 4.3 - Distribution of student answers for Item 3

Correct response = C

Choices with significant differences between animation and traditional = A

Table 4.3 - Summary of Statistical Data for Item 3

Item Three	Traditional	Animation
Correlation with English ACT	0.21	0.18
Correlation with Math ACT	0.26	0.42
% Correct	56	67
Index of discrimination	0.41	0.43

Points of Interest from the Interview Data

- Students who answered Item 3 correctly generally did so for the correct reasons.

For the most part, the students who correctly answered Item 3 did so because they understood that objects should speed up as they fall and that gravity is constant over short distances. They were often aware of terminal speed but correctly noticed that the building was one-story and stated that terminal speed should not have an effect over such a short distance. For example, the following statements where made by students who answered correctly.

> "Like on a one-story building its not that big a difference". [Referring to choice B] …"There were two, I think, that was constant [speed]. It didn't speed up [in those two] and it does speed up as it gets closer to the ground [so C is correct]" (Maggie, Animation)

> "I seem to remember something about terminal velocity. I'm not sure if it applies here … so I'll keep A in mind. B, I don't think that … would be relevant. That short distance wouldn't make that much of an effect. C, I think it might be C too because there is a constant force of gravity. I don't think D's the answer. I don't think E's the answer … I want to go with C, and it'll keep on speeding up because the gravity is constant." (Peter, Traditional)

- The animation group tended to debate between choices B and C, while the traditional group tended to debate between choices A and C.

This finding was very interesting and highly surprising to me as there was no obvious animation effect upon a first reading of the interviews. Five of the students who saw the animated question specifically stated that they had to decide between B and C. Two of the traditional students specifically stated that they were deciding between A and C. None of the other students appeared to be debating between two choices.

I went back through the transcripts and specifically focused on why the animation students easily discounted choice A and why the traditional group discounted choice B. In other words, why did each group discount the other groups' popular choice. It was not clear why the traditional group, as a whole, so quickly eliminated choice B. However, it does appear that seeing the animation of an object accelerating for a brief period of time and then abruptly falling at a constant velocity does discourage students from choice A. The following are statements made by students in the animation group for why choice A was incorrect.

> "That [Choice A] doesn't look right because it's kind of stopped half way and kept floating down. If you hit a terminal velocity you would go at the same speed all the way down from there. But I don't think one-story buildings big enough for that." (Lucy)

"I think it would keep accelerating, it wouldn't be the terminal velocity". (Nathan)

"That [choice C] looked the most natural, if I were watching it really happen, what I would pick." (Alexia)

The animation seems to steer the students away from choice A because the animation looks less natural to the students. The quantitative data supports the notion that the animation tended to steer students away from choice A. The traditional group was significantly more likely to answer A than the animation group in the large scale testing.

Did the animation have an effect?

As was discussed above, I believe the animation did have the effect of helping students to eliminate choice A because the motion depicted for choice A looked unnatural.

Recommendation - Traditional with Modifications

As I discussed in Section 3.1 I do not believe Item 3 is well written. However, the results of analysis do indicate that Item 3 could be a valuable question since most of the students who answered correctly did so for the correct reasons and the students who answered incorrectly displayed misunderstandings. This is reflected in the high level of discrimination for both the animated and the traditional versions.

Item 3 could be greatly improved by removing answer choices D and E and replacing them with choices that are consistent with the ideas being tested in choices A-C. Neither of these two answers were chosen by many students. In fact, the students often did not even know how to interpret these choices. For example Jake stated "and then D, I wouldn't know where to start on that." Perhaps D could be rewritten as, "Falls at a constant speed because of an almost constant force of gravity." And E could be, "Falls at a constant speed because of the combined effects of the force of gravity pushing it downward and the force of the air pushing it upward."

It is not clear if the animated or traditional version of Item 3 is superior. In one sense, the animated version appears better because it taps student's intuition. However, it may only be a matter of luck that their intuition leads them to the correct answer in this case. If, instead of a one-story building, the stone was dropped from an airplane then terminal velocity would be a factor and answer A would be correct. It is not clear to me that the students would have answered differently had the distance of fall been greater.

Since it is not clear which version is superior for assessment, I recommend the traditional version. It is easier to administer and requires less time for the students to answer.

4.3.4 Item Four

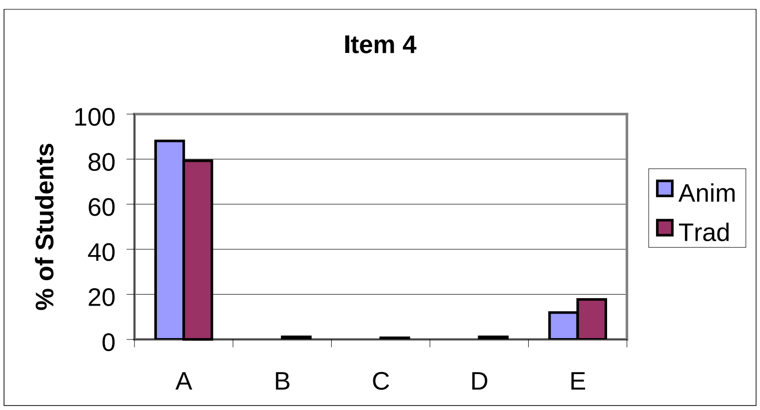


Figure 4.4 - Distribution of student answers for Item 4

Correct response = E

Choices with significant differences between animation and traditional = None

Table 4.4 - Summary of Statistical Data for Item 4

Item Four	Traditional	Animation
Correlation with English ACT	-0.03	0.03
Correlation with Math ACT	-0.03	0.09
% Correct	18	12
Index of discrimination	0.19	0.34

Points of Interest from the Interview Data

- Students consider information about how fast the car and the truck were moving to be important.

The question statement for Number 4 only says "A large truck collides head-on with a small compact car." It does not give any information about the speed of the car and the truck before or after the collision. Two students who saw the traditional version specifically asked me about speeds.

> "Are we assuming here that they're going the same velocity?" (Jason, Traditional)

> "Well, this is assuming they're both moving, right?" (Lucy, Traditional)

Two of the students who saw the animated version and answered incorrectly indicated speeds were important by using the animation to answer the question.

> "Well, the truck was pushing the car in the opposite direction that it was going, so the truck is going to exert a greater amount of force on the car." (Janice, Animation)

> "(Since) it goes in the direction that the truck was heading, I know that the truck exerts a greater amount of force. But it slowed down, so the car also exerted…some amount of force onto the truck" (Charlie, Animation)

There were two students, who originally saw the traditional version of the question and then later changed their answer when I showed them the animated version. These students are particularly interesting. Maggie actually answered Item 4 correctly the first time. Her reason was

> "I guess it kind of depends on, like how fast they were both going…If you had…a little car going really fast it can be as bad as a big truck going really slow. But they both got to end up pushing back the same amount, right?" (Maggie, Traditional)

When I asked her to look at the animated version of Item 4 she changed her answer, stating that the truck exerted more force than the car. She based her new answer on the fact that the truck was pushing the car out of the way.

Beverly gave answer choice D for the traditional version of the question. Her reason was

> "It seemed that the truck collided head-on with the car that they didn't actually meet somewhere. Say that the truck exerts a force on the car but the car doesn't exert a force on the truck." (Beverly, Traditional)

Upon viewing the animation she changed her answer to A and stated

> "The car didn't have that much of an impact on the truck, but it did hit it…The way I was reading it before, there was no picture, you said that the truck hit the car. And I wasn't for sure if the car was moving or not." (Beverly, Animated)

Apparently, the answer students give can be affected by whether or not the car and truck are moving before the collision and in which direction the pair goes after the collision. This information is present in the animation but is missing from the traditional version of the question.

- There is a common misconception that mass equals force.

Unsurprisingly, Item 4 did bring out the misconception that it was designed to test. Almost all students who answered incorrectly did so because they assumed heavier objects exert more force.

> "I believe since the truck is heavier it's going to have more momentum and more force" (Peter, Animation)

> "…because (the truck is) bigger, and I know that force has something to do with mass, so the bigger the mass the bigger the force" (Alexia, Traditional)

- The students who answer correctly tend to give Newton as a reason.

Only four students answered Item 4 correctly. Two of them specifically attributed their reason to Newton's 3^{rd} Law.

> "Alright, I'm thinking about one of Newton's laws that says for every action, there's an equal and opposite reaction." (Tanner, Animation)

> "…I believe that's one of Newton's principles, I can't remember it off the top of my head." (Nathan, Traditional)

This is an indication that there may be a large number of false positives associated with Item 4. It is possible that many students who answer correctly because they "remember" the correct response and not because they believe or understand the correct reasoning.

Did the animation have an effect?

I believe that the animation did have an effect in this question. The difference in the number of students answering A between the groups was significant at the $p = 0.05$ level, with the animation group being more likely to choose A. As I discussed above, this is probably because the animation showed both vehicles moving before the collision, and showed them both moving in the original direction of the truck after the collision. To many of the students, this was a clear indication that the truck must have exerted more force.

This question is very interesting. To the expert, dynamical information is not necessary to answer the question. Therefore, it would not be expected that an animation would affect the answers students give. However, dynamical information was essential to answering the question for many of the students. So in this case, the animation did have an effect.

Recommendation - Animation

Although Item 4 is not ideal due to the possibility of false positives, I do believe it gives a reasonable reflection of students' beliefs. The animated form seems to be a better reflection of students beliefs, because it is more specific in the question it asks. The traditional form could be rewritten to include a specific description of the initial and final conditions of the vehicles but this would add unnecessary information to the problem statement which may cause confusion. I recommend using the animated version of Item 4 because the information can be presented more simply in the animation.

4.3.5 Item Five

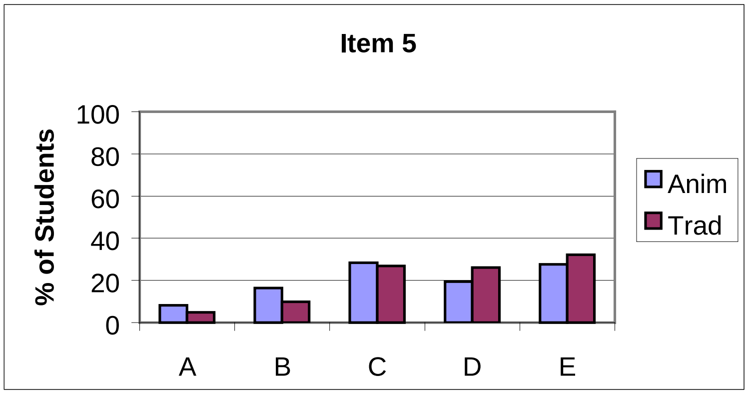


Figure 4.5 - Distribution of student answers for Item 5

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.5 - Summary of Statistical Data for Item 5

Item Five	Traditional	Animation
Correlation with English ACT	0.00	-0.08
Correlation with Math ACT	0.11	-0.05
% Correct	10	16
Index of discrimination	0.13	0.17

Points of Interest from the Interview Data

- Moving is associated with a force in the direction of motion.

There were no surprises in the way students answered Item 5. Almost everyone instantly and without hesitation included gravity as a force, even before reading the answer choices. Many correctly identified the radial normal force. Most incorrect answers were due to an inclusion of a force in the direction of motion. The most common reason for including the force in the direction of motion was because the ball was moving.

> "A force in the direction of motion. Yes, I would assume so because it has to be pushed forward for some reason. But it's a rapidly decreasing force." (Lucy, Traditional)
>
> "I guess if its moving its got to have that (a force in the direction of motion)." (Maggie, Traditional)
>
> "A force in the direction of motion because the ball's going to want to continue going that way." (Peter, Animation)

The answers students give to Item 5 do appear to be an accurate reflection of their understanding.

- Having Item 5 appear before Item 6 caused confusion on the animated version.

In the animated form of Item 5 students are asked about the forces on the ball in the channel. No picture is given. In order to determine the locations of the various points the student must activate one of the five animations that accompany Item 6. Note that Item 6 pertains to the same situation as Item 5 but the question involves the direction of travel of the ball after leaving the channel.

Several of the students were confused on Item 5 because it was not clear why there were five different animations. It would have been better to ask Item 6 before Item 5 on the animated version of the FCI or have an animation that stops at point r.

Did the animation have an effect?

There is no indication from either the statistical or interview data that the animation had any effect on the answer a student gave.

Recommendation - Traditional

Item 5 is a good question. Although the percentage of students answering correctly is lower than ideal, there is no indication of guessing. It appears that the answer a student gives is reflective of her/her understanding. Item 5 is not correlated with either verbal or math skills, which is good. It does have a low index of discrimination but that is probably due to the fact that so few students answered the question correctly. For example, only 10% of the traditional students gave the correct answer. The maximum discrimination would occur when all of the students answering correctly were in the "high" group. In this case that would mean a discrimination index around 0.22. Therefore the low discrimination found on this question is not a particular cause for concern.

There is no benefit to giving the animated version of the question over the traditional version. Since the traditional version is simpler to give, it is recommended.

4.3.6 Item Six

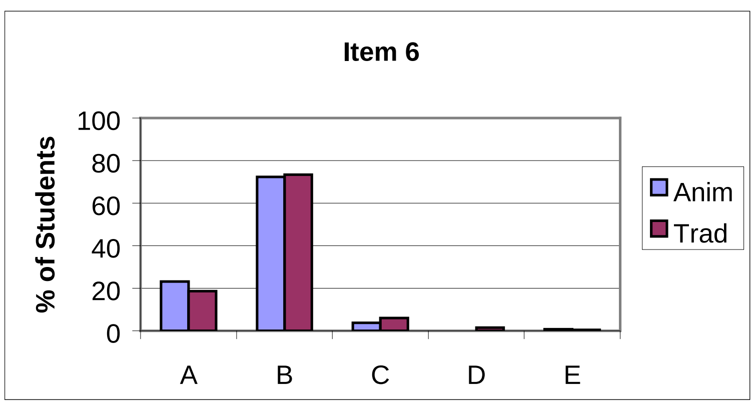


Figure 4.6 - Distribution of student answers for Item 6

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.6 - Summary of Statistical Data for Item 6

Item Six	Traditional	Animation
Correlation with English ACT	0.04	-0.05
Correlation with Math ACT	0.02	0.05
% Correct	73	72
Index of discrimination	0.27	0.20

Points of Interest from the Interview Data

- Item 6 is relatively uninteresting from an assessment standpoint.

All but two students answered Item 6 correctly. This is consistent with the high percentage of students answering correctly in the large scale data collection. For the most part, the students who answered correctly seemed to have a correct understanding. For example,

> "Because after you let it go that's not going to cause it to curve or anything." (Charlie, Animation)

> "It's not going to have anything to keep curving it anymore." (Jason, Traditional)

Both students who answered incorrectly chose answer A. Their reasons were

> "Because there is still going to be a force pointing toward the origin and its going to keep it going in that direction." (Janice, Animation)

> "A just … looks more right to me… I think it'd want to complete the circle." (Beverly, Traditional)

Both Janice and Beverly were asked to look at the opposite form of the question after completing the test. Neither felt she should change her answer.

Did the Animation have an Effect?

There is no indication from either the statistical or interview data that the animation had any effect. However, this is actually a very exciting result. In Chapter 3 of this dissertation, I reviewed a study by McCloskey and Kohl which was very similar to my study of Item 6. They also found no difference in the performance of the animated and the static groups.

An animation effect would be expected on this question since the question involves dynamical information and the animation is central to answering the question. No animation effect was seen. It is likely that an animation effect was not seen because this is a question that most students can answer correctly so there is little room for improvement. Interestingly, Item 6 is very similar to Item 7 and an effect was seen on Item 7. This issue will be discussed further in the discussion of Item 7.

Recommendation - Discard Entirely

Since most students answer Item 6 correctly, it is not particularly useful as an assessment item. This is especially true since these are pre-test results. Presumably, the proportion of students answering correctly on the post-test will be higher, thus lowering the value of the question even further.

Should Item 6 remain, there is no reason to use the animated version as the traditional version provides the same information in a simpler fashion.

4.3.7 Item Seven

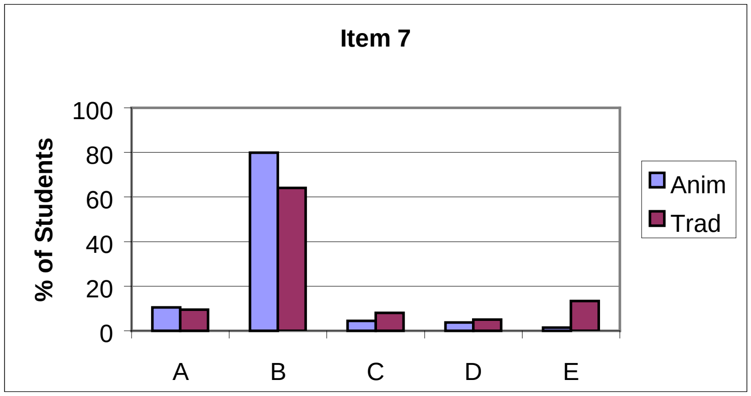


Figure 4.7 - Distribution of student answers for Item 7

Correct response = B

Choices with significant differences between animation and traditional = B, E

Table 4.7 - Summary of Statistical Data for Item 7

Item Seven	Traditional	Animation
Correlation with English ACT	0.05	0.02
Correlation with Math ACT	0.21	0.10
% Correct	64	80
Index of discrimination	0.35	0.23

Points of Interest from the Interview Data

- Many students specifically noted that Item 7 was almost identical to Item 6

No student indicated that he/she thought Item 7 was fundamentally different from Item 6. Many of them even referenced Item 6 as their basis of choice on Item 7. For example,

"I think it would be that one (B) for the same reason as the one before." (Lucy, Animation)

> "…because of the same thing I said in the above animation question." (Charlie, Traditional)

> "This will be a lot similar to the previous problem." (Nathan, Animation)

- The interview data is not consistent with the results of the large scale data collection.

Based on the large scale data collection, I would have expected about 65% of the traditional students to answer correctly and about 80% of the animation students. However, everyone eventually gave the correct answer. My interviewees did not fit the pattern seen in the larger sample.

Only Janice, who saw the traditional version, answered incorrectly on the first pass. She answered A because "the force is still going to keep it in that direction."   She had given the same answer choice and reason for Item 6. When I asked her to view the opposite forms of Items 6 and 7 she kept her answer to Item 6 without comment and changed to answer to Item 7 to the correct response with the comment "its not following the exact path, there is not anything holding on to it anymore, it would just go off in the direction it was let go."

I can not explain why I did not see the same pattern in my interviewees as was seen in the larger sample. This is particularly disappointing because the statistical data is very surprising and interesting. I saw essentially no differences in the number of correct responses on the two versions of Item 6, which was consistent with the results of McCloskey and Kohl. I would have expected the same result for Item 7 since it is essentially the same as Item 6. However, the animation group was statistically more likely to answer correctly for Item 7. Why would the students have responded differently to the two questions?

It is unlikely that the significant difference is due to random fluctuations. When McCloskey and Kohl investigated student responses regarding the path of a ball breaking from a rotating string they also found the animation group scored 15 percentage points higher than the traditional group. Given that their results are so consistent with mine, I feel that the effect is probably real, and not due to random variations. I was hoping that the interview data would help explain why the response pattern was different for these two questions but it did not clearly do so.

There are two possible explanations for why the interview data did not reflect the statistical data. First, all interviewees saw an animation on either Item 6 or Item 7. The statistical data is largely based on a population that saw either a fully animated or fully traditional FCI. It is possible that the results of the interviews were tainted by the students having essentially seen the opposite form of the question before answering Item 7. The fact that many students specifically noted that Item 7 was almost identical to Item 6 supports this. If there was a lingering animation effect from Item 6, that would explain why the interview results of Item 6 were consistent with the statistical results while Item 7 was not consistent.

Secondly, the think-aloud interviews required students to articulate their reasoning. This probably caused them to think about each question more thoroughly than they would have if they were taking the test in a normal fashion. I saw evidence of students changing their answer from a wrong response to a correct response as they articulated their reasoning to me. For example, Justin was one of the subjects that was allowed to work through the entire test before discussing answers with me. He saw the animated version of Item 7 and had originally answered E, incorrectly. When I asked him to explain his choice to be he replied…

> Justin: I kinda just looked at this one and I went with … what was it … E. It just seemed to me, (?) Now that I think about it I should have put C.

A little while later, he was back to choice E.

> Justin: … actually I think I'm going to go with E.
> M: So why, why E.
> Justin: I guess cause that one (Animation C) just looks really really awkward. Like, it wouldn't tend do that. That the ball would have more of a force, a force… I don't know it just seems like this would be a more natural motion.

And later on, he switches to B.

> Justin: Actually now B looks more natural than the others because when it comes around here its (?) motion is taking it more in that general direction, or its going to end up in that general direction. Where as with … C that kind of, that trajectory there, would have been at this point, this one would have been… (?) towards down here I think, I guess it'd be B but I'm not really sure.

By asking students to explain their reasons, I may have actually caused them to change their responses. As Justin demonstrated, students' ideas can be very fragile.

Did the Animation have an Effect?

The statistical data indicates that the animation had the effect of shifting students from the incorrect answer E to the correct answer B. This makes sense since the animation for E looks very unnatural. There was no obvious animation effect seen in the interviews.

Recommendation - Future Research

The pair of Items 6 and 7 turned out to be very complicated. My method of data collection demonstrated that both of these questions may contain very interesting animation effects. However, my methods did not clearly illuminate the inner workings of these effects. A concentration on just these two questions would be worthwhile for a future research project which will be discussed in Chapter 5.

The statistical data indicates that the usefulness of the question may be limited by the high percentage of correct responses, especially for the animated version. Although this question may be able to provide more information about the effect of animation, it may turn out that it is answered correctly too often to be valuable for assessment.

4.3.8 Item Eight

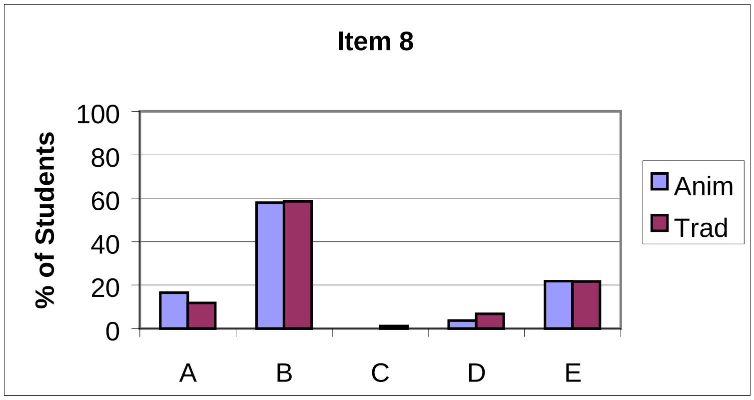


Figure 4.8 - Distribution of student answers for Item 8

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.8 - Summary of Statistical Data for Item 8

Item Eight	Traditional	Animation
Correlation with English ACT	0.06	0.01
Correlation with Math ACT	0.16	0.23
% Correct	59	58
Index of discrimination	0.38	0.54

Points of Interest from the Interview Data

- The answers students give are reflective of their thinking.

The students who answered correctly generally understood why path B was correct. For example,

> "…its still sliding that way so it'd be that (points right) plus that (points up) motion…that would put it on a constant velocity that way (right) as well as one that (up) way." (Nathan, Traditional)

> "…basically it's just going to be vector adding." (Jason, Traditional)

> "It's the combination of the force going that way and the force going that way." (Peter, Animation)

Note that Peter had the right idea but he described it in terms of force instead of motion. Several others also spoke of the original motion as a force to be added to the force of the kick.

All of the students who answered incorrectly gave A as their answer. There was no consistency among them for a reason for choice A. For example, Beverly saw the traditional version and interpreted the statement "Had the puck been at rest at point b then…" to mean that the puck slid to point b and then stopped before it was kicked. Interestingly, she did not change her answer when I later showed her the animated form of Item 8. Other reasons for choice A were…

> "I'm not seeing much of a difference between if it were sliding or if it had just been at rest…I don't think it would be B because there's no force to make it turn to the right." (Janice, Animation)

> "I think if the person who kicks the ball kicks it hard enough it's going to go straight." (Alexia, Traditional)

Did the Animation have an Effect?

I did not see any indications, from either the statistical or the interview data, that the animations affected the students answers

Recommendation - Traditional

Item 8 appears to be an excellent question. For both versions, the difficulty is near ideal, there is no correlation with verbal skills, and the correlation with math ability is weak. Since the animation does not appear to have any effect for this question, the traditional form is recommended for its simplicity.

4.3.9 Item Nine

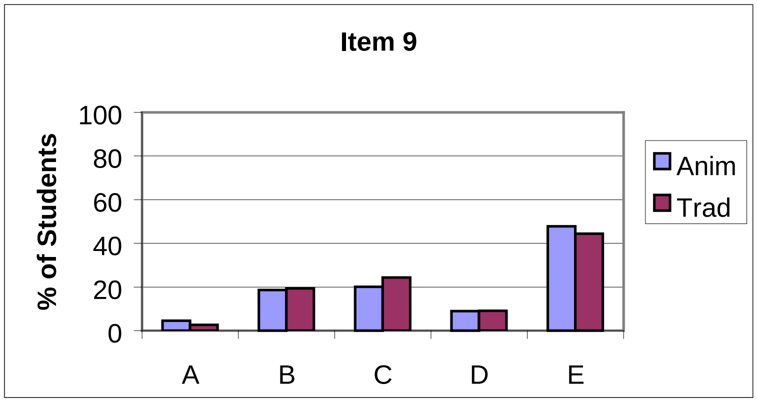


Figure 4.9 - Distribution of student answers for Item 9

Correct response = E

Choices with significant differences between animation and traditional = None

Table 4.9 - Summary of Statistical Data for Item 9

Item Nine	Traditional	Animation
Correlation with English ACT	0.24	0.31
Correlation with Math ACT	0.24	0.28
% Correct	44	48
Index of discrimination	0.25	0.40

Points of Interest from the Interview Data

- Students' answers are generally reflective of their understanding.

The students who answered correctly were able to demonstrate that they understood how to add vectors. For example,

"…because the speed, would've, the vector would have been the same as a Pythagorean triangle." (Charlie, Animation)

"I guess arithmetic sums just means 10 that way and 10 that way, so it'll be 20 which wouldn't be the case. It'd be the square root of two hundred, no try two thousand, no two hundred, that's right. " (Nathan, Traditional)

Reasons for incorrect answers varied but all incorrect answers reflected a poor understanding. Students who answered Item 8 incorrectly also tended to answer Item 9 incorrectly. For example, Tanner answered "A" for Item 8. For Item 9 he stated,

"Equal to the speed V_k resulting from the kick and independent of the speed V_o because it's in a different direction. I guess the horizontal velocity has nothing to do with the vertical velocity." (Tanner, Animation)

Did the Animation have an Effect?

There is no indication that student answers were affected by the presence or lack of an animation. The difference in the index of discrimination between the two versions was large enough to take note of, but given a complete lack of evidence to support a true difference, I believe it was due to statistical chance.

Recommendation - Traditional

Item 9 appears to be an accurate reflection of student thinking. And as with other questions where no animation effect was seen, the traditional version is recommended for simplicity.

4.3.10 Item Ten

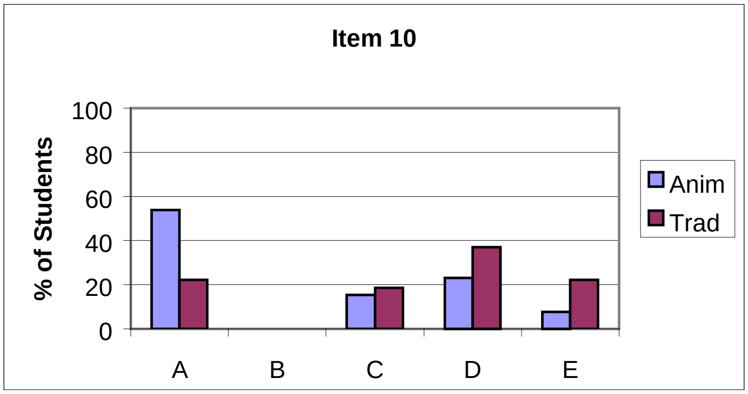


Figure 4.10 - Distribution of student answers for Item 10, Note that distribution is only based on Sample 1 (N=53) for this question due to a problem with the data collection. Due to the small data sample, statistical significance is difficult to achieve.

Correct response = A

Choices with significant differences between animation and traditional = None

Table 4.10 - Summary of Statistical Data for Item 10. For this table, % Correct is only based on Sample 1, other data is only based on Sample 2. Note that no data was collected for the animation group for Sample 2 for this question.

Item Ten	Traditional	Animation
Correlation with English ACT	0.09	N/A
Correlation with Math ACT	0.20	N/A
% Correct	22	54
Index of discrimination	0.35	N/A

Points of Interest from the Interview Data

- A correct answer reflects a correct understanding.

All interview students who answered Item 10 correctly stated that the speed remains constant because there are no outside forces. For example,

> "…if you give it an initial velocity and there is no other force acting on it then it could just keep going forever." (Janice, Animation)

> "I think the speed would have to be constant because there is no friction. Nothing to resist it." (Georgia, Traditional)

- There is a common belief that everything **must** stop moving.

Without exception, students who answered incorrectly indicated that they believed that the puck must eventually stop. Two students attributed the stopping to gravity, one attributed it to air resistance and the other two simply said it must stop. For example,

> "…since gravity is constant it would be constantly acting on it and decrease the speed of the puck." (Justin, Traditional)

> "It's already received the kick so all it has left is to get rid of energy and decrease…even though this one doesn't have friction it still has air resistance, and it should just continually decrease rather than be constant." (Barb, Animation)

> "Well, it's frictionless but in some way it's going to have to stop…it has to sometime decrease." (Lucy, Traditional)

Did the Animation have an Effect?

Only one animation student answered incorrectly while half of the traditional students answered incorrectly. This result is consistent with the statistical result. It should be noted that the main reason a statistically significant result was not reported from the large scale data collection is because only 26 students answered the animation question and 27 answered the traditional question. Although the difference in the proportion of students answering correctly on the two forms was large, the small sample size makes it difficult to claim the difference was significant.

Since the spread of answers in the interviews was consistent with the large sample data, I do believe the animation had the effect of steering students towards the correct response. However, I did not see any clear indications in the interview data that could offer an explanation.

When I first wrote the animated form of Item 10, I believed that the animation would help students answer correctly. After all, they could simply view the animation to get information about speed and did not have to use their own understanding as a basis. However, I saw no indication that this was happening. No student looked back at the animation when answering Item 10. Also, without exception, students gave a correct reason based in physics when they answered correctly.

The most likely explanation for an animation effect is that the students were subconsciously using the speed information from the animation. None of the animations depicted the speed of the puck decreasing. It is possible that this visualization imprinted on the students in a way which they were not aware and could therefore not articulate.

Recommendation - Traditional

Although there does seem to be some animation effect on Item 10, the animated version is not clearly superior. Also, there are no indications that the traditional version is problematic from an assessment standpoint. Therefore, I recommend using Item 10 in its traditional form. Also, since Item 10 goes with Items 8, 9, and 11 and the traditional form is recommended for those questions, the traditional form of Item 10 is recommended for consistency.

4.3.11 Item Eleven

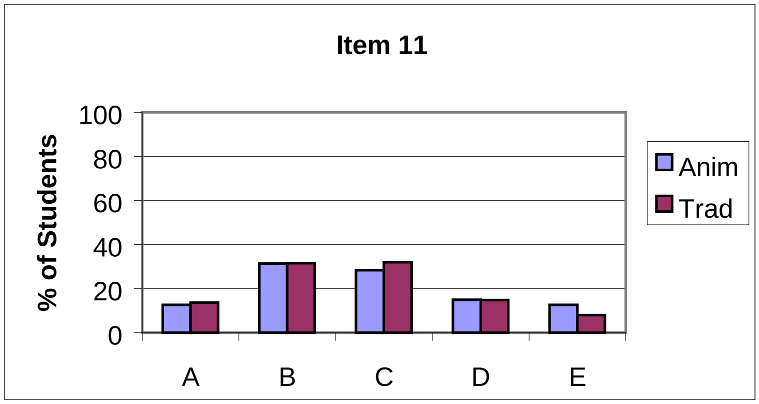


Figure 4.11 - Distribution of student answers for Item 11

Correct response = D

Choices with significant differences between animation and traditional = None

Table 4.11 - Summary of Statistical Data for Item 11

Item Eleven	Traditional	Animation
Correlation with English ACT	0.11	0.06
Correlation with Math ACT	0.09	0.05
% Correct	15	15
Index of discrimination	0.24	0.29

Points of Interest from the Interview Data

- There is a common misconception that an object in motion must have a force acting on it.

The results of Item 11 were very similar to those of Item 5. A large number of students believe that there is force acting on an object in the direction of its motion. Most students do not even justify this force, they simply state it as if it were obvious. When they do give a reason it tends to be along the lines of, "if it's moving, something else has got to be pushing it" (Maggie, Traditional). It did not seem to bother Maggie that the "something else" was completely undefined.

- Students can incorrectly identify the upward force of the surface as friction instead of a normal force.

Some students interpreted "an upward force exerted by the surface" to mean friction. These students knew that friction was not present so they eliminated choices with this force. For example,

> "There's no upward force exerted by the surface because we were just assuming there's no friction." (Jason, Traditional)

Did the Animation have an Effect?

There is no indication that the animation had any effect.

Recommendation - Traditional

Since the animation had no effect the traditional form of Item 11 is recommended for simplicity.

4.3.12 Item Twelve

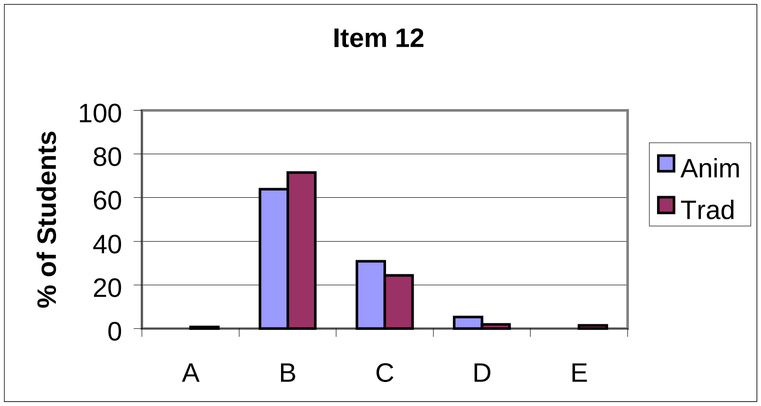


Figure 4.12 - Distribution of student answers for Item 12

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.12 - Summary of Statistical Data for Item 12

Item Twelve	Traditional	Animation
Correlation with English ACT	0.03	-0.05
Correlation with Math ACT	0.05	0.03
% Correct	71	64
Index of discrimination	0.40	0.34

Points of Interest from the Interview Data

- Most students understand the concept behind Item 12

Only one interview student answered Item 12 incorrectly. Most of the students who answered correctly were able to demonstrate that they understood why B represented the correct path. For example,

> "…because as soon as it leaves the cannon it's being affected by gravity." (Tanner, Animation)

> "It's going to be constantly falling…it keeps on accelerating…I don't think it goes straight any…It doesn't go straight for an amount of time because it's accelerating downward at all points." (Lucy, Traditional)

However, there were three students who gave the correct answer but demonstrated misunderstandings. Charlie (animation) spoke of the ball being "affected on both by the force of being shot and gravity acting on it" demonstrating the force in the direction of motion misconception. Maggie (traditional) stated that it might be choice A if "it was really, really fast." And Janice (animation) only discounted C because "it went straight for too long" not because it went straight at all.

Only Alexia (traditional) answered incorrectly. She debated between choices B and C and finally decided on choice C because "it's going to be propelled straight before it falls." When I later asked her to look at the animated form of Item 12 she kept her answer as C. She stating that "the force exerted by the cannon is going to be greater than gravity and it's not going to affect it until the ball starts losing, till that force dissipates and it starts being overcome by gravity."

Incidentally, one common concern with this question is that students do not notice the difference between choices B and C. It should be noted that I saw no indication of this difficulty in either version during the interviews.

Did the Animation have an Effect?

There is no indication that the animation had an effect.

Recommendation - Traditional with Modifications

Since the animation had no effect the traditional version is recommended. Most students who answered Item 12 correctly did so for the right reasons, but not all. I believe some of these false positives could be eliminated by altering the question situation so that the projected object was "faster". For example, the object could be a bullet from a high-powered rifle. One weakness of Item 12 is that so many students answer correctly. The modifications would probably decrease the number of correct responses and thereby improve the question all around.

4.3.13 Item Thirteen

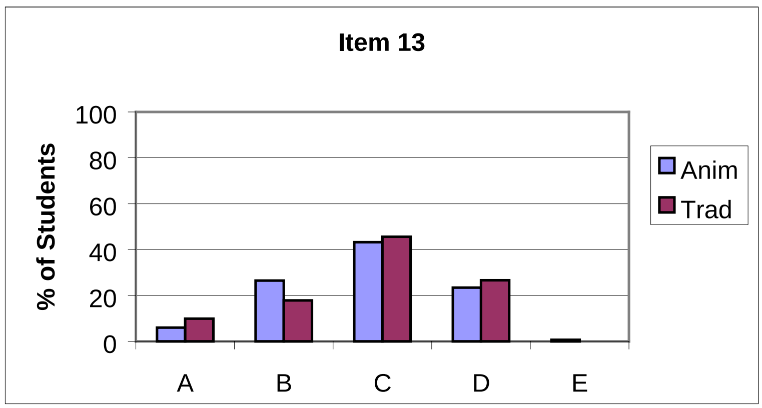


Figure 4.13 - Distribution of student answers for Item 13

Correct response = D

Choices with significant differences between animation and traditional = None

Table 4.13 - Summary of Statistical Data for Item 13

Item Thirteen	Traditional	Animation
Correlation with English ACT	-0.07	0.13
Correlation with Math ACT	0.01	0.35
% Correct	27	23
Index of discrimination	0.41	0.57

Points of Interest from the Interview Data

- Student ideas are inconsistent.

Very few of the interviewees answered Item 13 correctly. It is difficult to explain the thought processes of students as they answered this question because they applied their ideas inconsistently. I saw several misconceptions.

1. Force = Velocity
2. There is a force in the direction of the motion.
3. The force of gravity only affects falling (not rising) objects.

However, these misconceptions were not clearly and consistently applied by each student. For example, consider Peter who answered C.

> "I believe [in] the very beginning you have gravity acting on it to slow it down…so the upward force decreases the moment it leaves the boy's hand. Then on the way down, gravity pulls it down." (Peter, Animation)

Peter seems to believe that there is a force in the direction of the ball's motion on the way up which decreases as the speed decreases. However, he does not continue this concept as the ball falls stating that there is only the constant force of gravity on the way down. Lucy, who also chose C, demonstrates the same inconsistent understanding.

> "…there is an upward force…not only does it steadily decrease but at some point becomes zero because it's going to have to go back to the ground…[gravity is constant on the way down because] the mass acceleration never changes so the force of the Earth does not increase." (Lucy, Traditional)

Lucy also seems to be equating force and velocity as the ball rises but not as it falls. Jason also chose C and shows inconsistencies in his response.

> "There's going to be the force that he's going to have to throw it up at the beginning because, I mean, otherwise there's going to be no way it can get up." (Jason, Traditional)

Like the others, Jason does not clearly indicate why the decreasing upward force applies on the way up but not on the way down. Unfortunately, this inconsistency was not as obvious to me as I was doing the interviews as it is now. So, I did not specifically ask any of the students to explain. I did ask Georgia, who also answered C, "What is the upward force?" She answered, "I don't know…I guess it wouldn't be a normal force, that's what I was thinking. But we read that all forces have an opposing force." It appeared to me that she believed there should be an upward force and was willing to stretch her ideas in order to accommodate it.

- The animation was not always viewed.

Two of the animation students answered the question correctly. One of them never viewed the animation and the other only viewed it after giving his response. They apparently did not feel viewing the animation was necessary to answer the question.

Did the Animation have an Effect?

There is no indication that the animation had an effect. The animated version is slightly more correlated with math skills than the traditional version. However, due to a lack of any theory or support for the difference in the data, I believe it is just a statistical fluctuation and not a real difference.

Recommendation - Traditional

Item 13 appears to be an excellent question. It has a high level of discrimination and there are indications that the answers students give are reflective of their thinking. The animation is superficial and of no benefit. The traditional version of Item 13 is recommended.

4.3.14 Item Fourteen

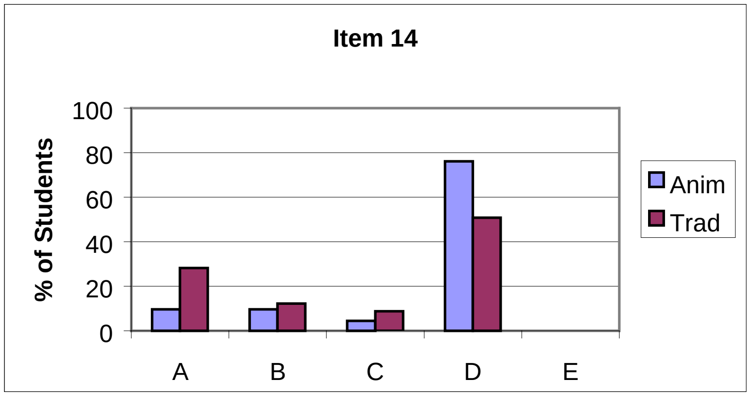


Figure 4.14 - Distribution of student answers for Item 14

Correct response = D

Choices with significant differences between animation and traditional = A, D

Table 4.14 - Summary of Statistical Data for Item 14

Item Fourteen	Traditional	Animation
Correlation with English ACT	0.15	0.12
Correlation with Math ACT	0.27	0.23
% Correct	51	76
Index of discrimination	0.56	0.37

Points of Interest from the Interview Data

- Many students correctly understand the concepts involved in Item 14.

Many of the interview students answered correctly and were able to demonstrate that they understood why D was the correct answer. For example,

> "…the ball's still going to have its horizontal velocity and gravity is slowly going to have … an effect upon vertical velocity." (Jake, Traditional)

> "Some things I can tell you right off before I even look at the animations is the bowling ball is going to hold the same velocity as the airplane so basically what I'm going to be looking for is something that's going to have a path like that. [draws a path like choice D]." (Jason, Animation)

However, two students who answered correctly also demonstrated misconceptions. Lucy (animation) believed that "the velocity forward is slowly decreasing" and Alexia (animation) stated that "[the ball's] going to use a little bit of [the plane's] motion until it's overcome by gravity and falls."

- Choice E is unnecessary.

No one chose E in either the interviews or the large data collection. Several of the interview students commented that E was very strange. Beverly, Alexia, and Maggie openly laughed after viewing Animation E. Others viewed the animation and commented,

> "No, it's definitely not E although that's quite humorous." (Lucy)

> "That one was super-glued to the bottom of the airplane." (Jason)

None of the traditional students commented about E at all. They tended to justify why all the other choices were correct or incorrect but ignored E. This is an indication that they thought E was so absurd that it did not need to be justified as an incorrect answer.

- Perspective has an effect.

There were two students who answered incorrectly primarily because they chose the path with respect to the plane instead of with respect to the ground. Georgia (animation) chose A and stated "It will tend to fall from where it was released and will fall backwards in respect to the plane." She clearly misunderstood the question being asked. Barb (traditional) also chose A. She admitted that she had a difficult time choosing an answer but finally settled on A because,

> "…you see a lot of movies where they have jumpers and…they always…you're seeing the plane in a constant position but…in movies…the plane's already moving on and so you just see…behind you. And so it's kind of hard to imagine what it's going to do exactly."

Barb seemed to be aware that her answer was based on the wrong perspective but could not translate her understanding of one perspective to another.

Did the Animation have an Effect?

The statistical data does indicate that the animation had the effect of shifting some students from choice A to the correct response D. There are indications from the interview data that some students who choose A do so because they are thinking about the path relative to the plane rather than relative to the ground. Therefore, it seems logical that the animation would reduce the proportion of students giving A as an answer and increase the proportion of students giving the correct answer.

Although the animation does not eliminate the perceptual problem it probably reduces it. I believe that the students who are most likely to be affected by the animation are students who have a correct understanding but lacked an attention to the detail of the question statement.

Recommendation - Animation with Modifications

I recommend the animated form of Item 14 because it appears to eliminate some false negatives due to perceptual issues. However, I must admit that it could be argued that the traditional form of the question is superior. My argument that those who may be shifted from an incorrect to a correct answer by the animation is only moderately supported by the interview data. The traditional version had a much higher index of discrimination which would indicate that the animation helped the weaker students. I have chosen the animated form of this question as the superior form because I feel the argument for the animated form is stronger than the argument for the traditional form.

In either version, I believe the question could be improved by replacing choice E with a path similar to choice C from Item 12. This path would be consistent with the idea that the ball will go some distance before gravity has an effect. Given that choice C was the second most popular choice for Item 12, it would probably also be popular if added to Item 14. And there would be no loss by excluding choice E. This change would also tend to decrease the number of correct responses which would be an improvement to this question since the proportion of correct responses is slightly high.

4.3.15 Item Fifteen

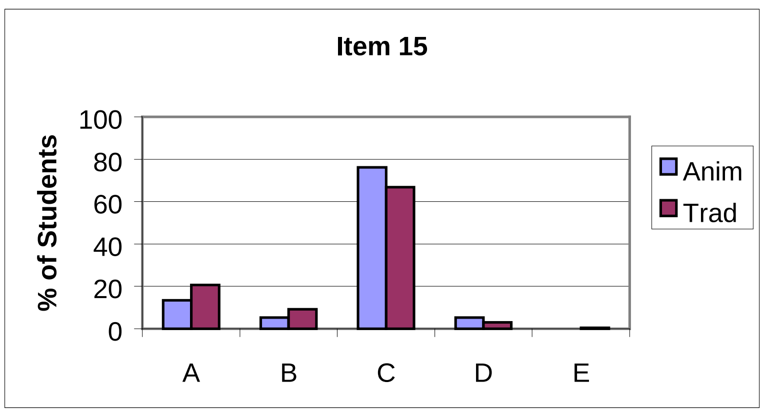


Figure 4.15 - Distribution of student answers for Item 15

Correct response = A

Choices with significant differences between animation and traditional = None

Table 4.15 - Summary of Statistical Data for Item 15

Item Fifteen	Traditional	Animation
Correlation with English ACT	-0.03	-0.05
Correlation with Math ACT	-0.09	-0.01
% Correct	21	13
Index of discrimination	0.15	0.31

Points of Interest from the Interview Data

- Item 15 brings out common misconceptions.

Only one of my interview students answered correctly. Most of the interviewees chose C for their answer. They all basically gave the same explanation for their choice. For example,

> "If [the force of the car equaled the force of the truck] then they wouldn't be going anywhere. And if the force of the car on the truck is smaller than the truck [on the car], then the truck would be pushing the car backwards." (Maggie, Animation)

> "Well A wouldn't be true or…neither of them would move so I'm looking for something where the car is pushing more than the truck." (Jake, Traditional)

> "If [the forces on the car and the truck] were equal they would not be going anywhere…if [the truck exerted a greater force on the car] then they would be moving the other way." (Lucy, Animation)

There was a common misconception that equal internal forces correspond to no motion and that there is a greater internal force in the direction of the motion. I also saw evidence that some students believe that only "active" agents can exert a force. For example, two students gave D as an answer. They both said that the truck exerts no force since it was not the vehicle doing the pushing.

Did the Animation have an Effect?

There is no indication that the animation had any effect.

Recommendation - Traditional with Modifications

Item 15 appears to be a good question. It is not correlated with either verbal or math skills and the discrimination is adequate. Although the proportion of students answering correctly is very low, I saw no evidence that there was any random guessing on the part of the students. The students understand the question being asked and their answers reflect their understanding. Since the animation had no effect the traditional version of Item 15 is recommended.

Answer choices A, B, and C should have the statement "Neither force is zero" added. The reason for this addition is discussed below in conjunction with Item 16.

4.3.16 Item Sixteen

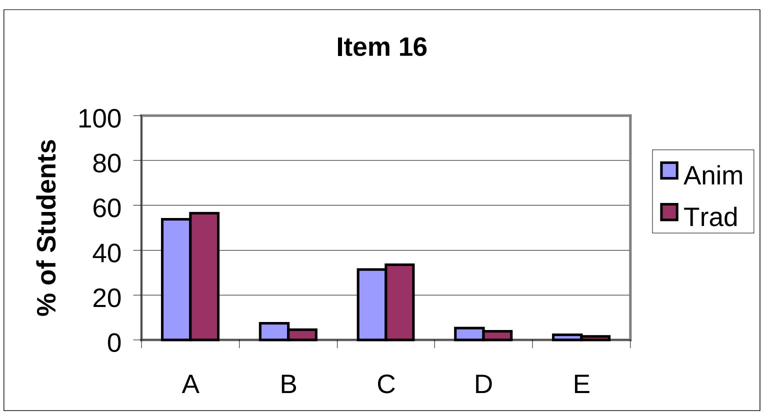


Figure 4.16 - Distribution of student answers for Item 16

Correct response = A

Choices with significant differences between animation and traditional = None

Table 4.16 - Summary of Statistical Data for Item 16

Item Sixteen	Traditional	Animation
Correlation with English ACT	-0.01	0.01
Correlation with Math ACT	0.01	0.17
% Correct	56	54
Index of discrimination	0.40	0.49

Points of Interest from the Interview Data

- Students who answered Item 15 incorrectly often answer Item 16 correctly because the car and truck are not accelerating.

A number of students who had given an incorrect answer to Item 15 answer Item 16 correctly. The reason they typically give for the difference in answers to the two questions is that the car and the truck are at a constant speed in Item 16 and accelerating in Item 15. For example,

> "There's no longer acceleration so the two forces would be equal." (Nathan, Animation)

> "…since they are remaining at the same velocity…they would not be accelerating or decelerating…I think that they would be equal." (Jake, Traditional)

These students are giving the correct answer, but the correct answer does not demonstrate that they understand the concepts involved.

As another example of a false positive consider Charlie (Traditional). Charlie answered Item 16 correctly but stated that "I think the forces on each other are going to be equal because they won't be acting on each other anymore." Charlie's response actually corresponds to choice E - that both forces are zero. However if both forces are zero, they are also equal so choice C also fit Charlie's ideas. Items 15 and 16 should be modified to distinguish the answer choices.

Did the Animation have an Effect?

There is no indication that the animation had any effect.

Recommendation - Traditional with Modifications

Since the animation had no effect, the traditional version of Item 16 should be used as it is easier to deliver. Answer choices A, B and C should be modified, as suggested in the recommendation for Item 15, to make it clear that the forces are considered to be non-zero.

It is interesting to note that Item 16 had higher indices of discrimination than Item 15. This is somewhat surprising since there are a number of false positives associated with Item 16. The lower index of discrimination seen on Item 15 is probably a result of the low proportion of students answering Item 15 correctly. As has been previously discussed, the index of discrimination is limited when the proportion of correct answers is very high or very low.

Since the index of discrimination found for Item 16 is excellent, my concern about the number of false positives can be overlooked. I believe that the combination of Items 15 and 16 can be very insightful for assessment. Those students who answer both questions correctly have the strongest understanding. Those who answer only Item 16 correctly have a partial understanding, and those who answer both incorrectly have a poor understanding.

4.3.17 Item Seventeen

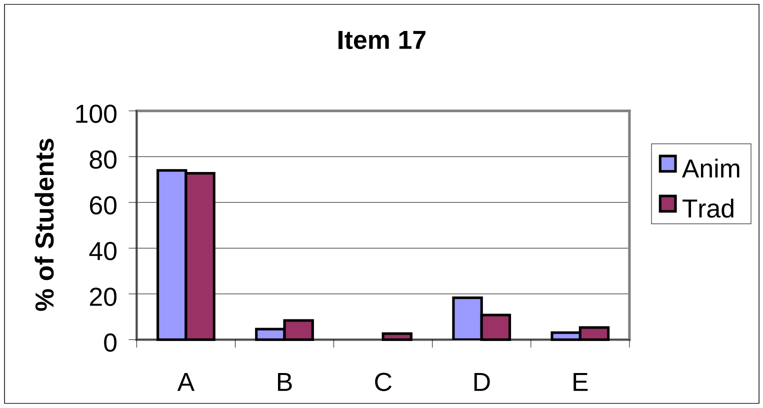


Figure 4.17 - Distribution of student answers for Item 17

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.17 - Summary of Statistical Data for Item 17

Item Seventeen	Traditional	Animation
Correlation with English ACT	-0.23	-0.20
Correlation with Math ACT	-0.26	0.04
% Correct	8	5
Index of discrimination	0.10	0.06

Points of Interest from the Interview Data

- Students believe there is a net force in the direction of motion.

All but one interview student answered Item 17 incorrectly. The student who answered correctly noted that the forces were equal because there was no acceleration. Without exception, the other students all stated that the force up was greater than the force down because the elevator was moving up. For example,

> "I know that the force that the elevator is being pulled up has to be greater than gravity because it is going up." (Jake, Traditional)

> "It is going up so the cable must be pulling harder than the gravity." (Maggie, Animation)

Did the Animation have an Effect?

There is no evidence to suggest that the animation had any effect. In fact, the animation was not even activated by all of the students who were presented with the animated version of the question. Apparently, they did not feel the animation was needed to answer the question.

Recommendation - Traditional

Since there is no animation effect the traditional version is preferred over the animated version. Also, the answers students give do seem to be reflective of their understanding. However, there are several problematic issues concerning Item 17.

First, there is a negative correlation between both math and verbal skills for this question. This negative correlation is odd and could indicate a fundamental problem but is most likely due to the extremely small number of students answering the question correctly. Secondly, the index of discrimination is almost zero, making this question worthless for distinguishing among students. However, I believe that both of these findings can be accounted for by the extraordinarily low number of students answering correctly. Because so few students answered correctly the index of discrimination is severely limited and the correlations are based on a very small sample.

Ordinarily, an assessment question answered correctly by only 8% of students would not be considered a valuable addition to a test. However, these results are based on a pre-test and there is no indication of guessing. Also, the goal of conceptual assessment is not always to assess students. Sometimes it is meant as an assessment of a class or a curriculum. If that is the case then an item which is commonly answered incorrectly raises no difficulties.

Since Item 17 does appear to be a valid measure, and since the animation is superficial, the traditional form of Item 17 is recommended.

4.3.18 Item Eighteen

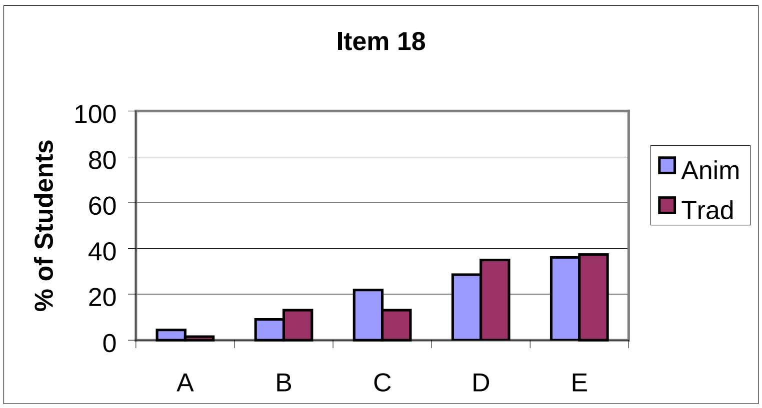


Figure 4.18 - Distribution of student answers for Item 18

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.18 - Summary of Statistical Data for Item 18

Item Eighteen	Traditional	Animation
Correlation with English ACT	0.01	0.02
Correlation with Math ACT	-0.02	0.10
% Correct	13	9
Index of discrimination	0.18	0.14

Points of Interest from the Interview Data

- Students believe there must be a force in the direction of any motion.

Unsurprisingly, all but on interview student included the "force in the direction of the boy's motion" as a force acting on the boy in the same way as they did for Item 5. They believe there must be a force in the direction of motion because the boy is moving that direction. For example,

> "There's probably a force in the direction of the boy's motion because he is going back and forth." (Tanner, Traditional)

> "A force in the direction of the boy's motion, just because he is the force, well he's part of the force and he's in motion." (Alexia, Animation)

- The answers students wanted to give were not necessarily a choice.

Two students found that the combination of forces they believed to be correct was not an answer choice. Lucy felt that the force from A to O and the force from O to A should be included. None of the choices included both of these forces. Georgia felt that the forces would be gravity and the force from O to A only, but this was not a choice. Since their preferred choice was not available, both women gave a response they did not feel was correct.

For these two students, the response they gave was not indicative of their understanding. This represents a shortcoming of this question. However, there can only be five answer choices. Not all combinations can be represented. According to the authors of the FCI, the answer choices are based on the most common open-ended answers given by students. Presumably, these five choices represent the largest number of student ideas.

Did the Animation have an Effect?

There is no indication that the animation had any effect.

Recommendation - Traditional

As with all the other questions where no animation effect was seen, the traditional version of Item 18 is recommended. With the few exceptions noted above, the answers students give do seem to be an accurate reflection of their understanding. There is little correlation with math or verbal skills and the index of discrimination is fine, especially since the proportion of correct answers is low. As with Item 17, the proportion of students answering correctly is low, but probably not of concern for the same reasons discussed before.

4.3.19 Item Nineteen

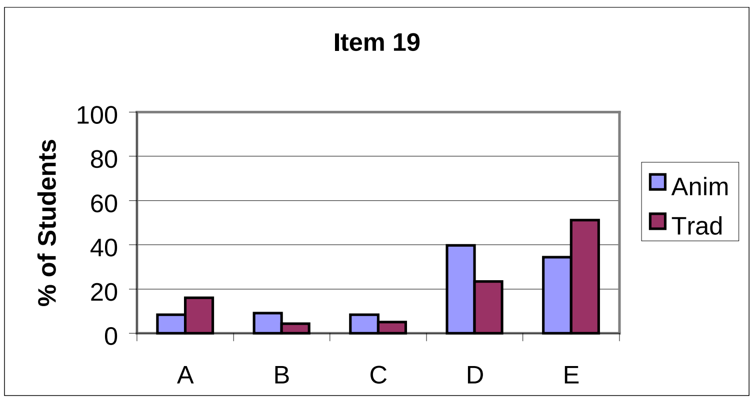


Figure 4.19a - Distribution of student answers for Item 19

Correct response = E

Choices with significant differences between animation and traditional = D, E

Table 4.19 - Summary of Statistical Data for Item 19

Item Nineteen	Traditional	Animation
Correlation with English ACT	0.27	0.30
Correlation with Math ACT	0.25	0.29
% Correct	51	34
Index of discrimination	0.41	0.60

Points of Interest from the Interview Data

- When students answer correctly, they are generally able to give a correct explanation.

Almost all of the students who answered Item 19 correctly explained that the speeds of the two blocks were the same between 3 and 4 because the blocks had traveled the same distance in the same time. This was true for both the animation and the traditional students. For example,

> "They both increased in the same intervals, per second, or per fraction of a second." (Barb, Animation)

> "I just noticed that from 3 to 4 they seemed to increase the same…the distance between them is the same…for the top block and the bottom block." (Georgia, Traditional)

- Some students who answer incorrectly equate position with velocity.

Although very few of my interview students answered Item 19 incorrectly, the most common incorrect response was D. When students choose D they are implying that the blocks have the same speed when they are at the same position. For example, Barb (who is also quoted above) originally chose D because "that was what stuck out to me because of the places where they met up, but then I realized that was the same position, not the same speed." Jason (Traditional) chose D because "they have the same *x* at the same *t*."

- Students who see the traditional version may not understand what the diagram represents.

Two students who saw the traditional version of Item 19 could not interpret the diagram given. After reading the question statement, Beverly said, "Uh, I don't think I understand this question." She tried to make sense of it but was unsuccessful.

> "I'm trying to figure out if, when it's saying that t equal one second if, I don't know what I'm thinking. Uh, is it the one second before it comes there, or the one second after the, like when it's going to the 2 is what I'm thinking." (Beverly, Traditional)

She eventually concluded,

> "I honestly have no idea what this question is asking. Or I know what it's asking, but I'm at a loss of how to figure this out so I'm going to try to take an educated guess. And my guess is, [long pause], um, at 5, C. (Beverly, Traditional)

I later asked her to look at the animated version. Although she did not answer correctly, she did appear to understand what was being asked of her.

> "On that one [the animation], it all looks different. Like the blue one is…[replays animation]. I'm going to change that one to D…Because that time they look like they were going the same speed at one and four." (Beverly, Animation)

Justin had similar problems. He also saw the traditional version of the question. Justin was one of the students who worked through the entire test and then discussed answers with me. He began our discussion by asking me to clarify what the diagram represented.

> "This one, I wasn't real sure on. [?] should that be just 0.2 second intervals, which would mean that would be 1.6 seconds?" (Justin, Traditional)

I answered his question and gave a short explanation of the diagram. But it did not seem to help him.

> "My answer was A and I really had a hard time visualizing and thinking about what all it entailed, and to come up with the right answer. I was looking at E because it seemed like you wouldn't have enough information to pinpoint one exact time. But, at E, between two and three seconds this block comes in with a higher speed and maintains a higher speed, this one comes in with a lower speed and maintains that lower speed. Though I guess it could fluctuate between, there was a little uncertainty in my answer." (Justin, Traditional)

Although Justin understood what was being asked of him, he was not able to "read" the diagram well enough to answer the question. I interviewed Justin before I started having students go back and look at the opposite form of some questions. So I do not know how Justin would have responded if he had later been shown the animation.

- The interview results are not consistent with the large-scale statistical data.

The results of the large data collection indicate that the traditional group was significantly more likely to answer Item 19 correctly than the students who saw the animation. The exact opposite was true of my interviewees. All of the students who saw the animation answered Item 19 correctly. Three of the eight traditional students answered incorrectly. I can not explain this discrepancy, though it may simply be an artifact of the small number of interviewees.

The interview data clearly indicates that students are more likely to answer the traditional version of the question incorrectly because they do not understand what the diagram represents. Common sense also supports this finding. It seems logical that the animation question would be easier for students because they can rely on their intuitions more and less on their ability to makes sense of an abstract depiction.

The result of the statistical data collection is surprising and needs to be explained. Student data on this question was collected as part of a pilot study for this dissertation. That data does not explain the discrepancy between the interview data and the statistical data collected as part of the current study. However, it does add some extremely interesting insights.

As part of the pilot study, I collected student responses to a slightly different version of Item 19 from four different groups. These groups saw the blocks moving across the screen without leaving the footprints seen by the students in the current study. All students in the pilot study answered Item 19 post instruction. Below is a brief summary of each group.

University 1 - (N=45) This group answered both the animated and the traditional form of Item 19 as part of their final exam. The students where taking calculus-based mechanics at a large state university. They were part of an experimental, highly interactive, classroom.

University 2 - (N=102) The students in this group were randomly assigned to either the animation or traditional condition. They answered the questions outside of class and were not graded on correctness, only completion. They were taking a lecture style calculus-based mechanics at a historically black state university.

College - (N=30) The students in this group answered both the traditional and the animated version. They answered the two versions several days apart. The students were randomly assigned to groups such that half the students saw the traditional version first and half saw the animated version first. The students were taking a lecture style algebra based mechanics course at a highly selective private liberal arts college.

High School - (N=18) The students in this group were randomly assigned to either the animation or traditional condition. They were part of a conceptual physics course at an affluent private high school and answered Item 19 as part of their final exam.

The results of the pilot study are shown in Figure 4.19b graphs below.

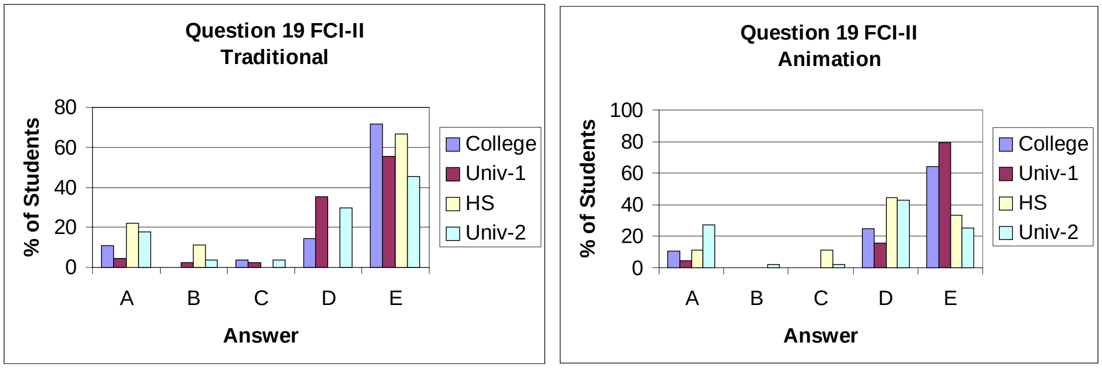


Figure 4.19b - Distribution of students' answers to Item 19 during the pilot study.

It is interesting to note that the University 1 students performed significantly better ($p < 0.05$) when they saw the animation than they did when they saw the traditional version. There was no significant difference in performance on the two versions for the College and the High School groups. And the University 2 students performed significantly worse ($p < 0.05$) on the animated version compared to the traditional version. Note that the High School animation group performed 30% worse than the animation group but due to the small number of students, this difference was not significant.

So, the results of the pilot study are just as odd and discrepant as the results of the current project. I can not explain why some groups seem to be aided by the animation and some seem to be hindered. The effect of the animation probably depends on the background and abilities of the students. The groups in the pilot study were very different from each other in academic background, socioeconomic status, and instructional style.

Did the Animation have an Effect?

The animation does appear to have an effect on the answers students give to Item 19. However, as I discussed above, both the direction and the nature of this effect are unclear. Based on the interview data, I believe the animation can reduce confusion about the question statement.

Recommendation - Animation

I believe that the animated version of Item 19 is superior because it is more valid. As I discussed above, some students may essentially guess at the traditional version of this question since they are unable to interpret the meaning of the diagram. Although the animation may not help students answer correctly, I believe it provides a better assessment of their understanding. It should also be noted that the index of discrimination for Item 19 is exceptionally high, making this an excellent question for comparing students.

4.3.20 Item Twenty

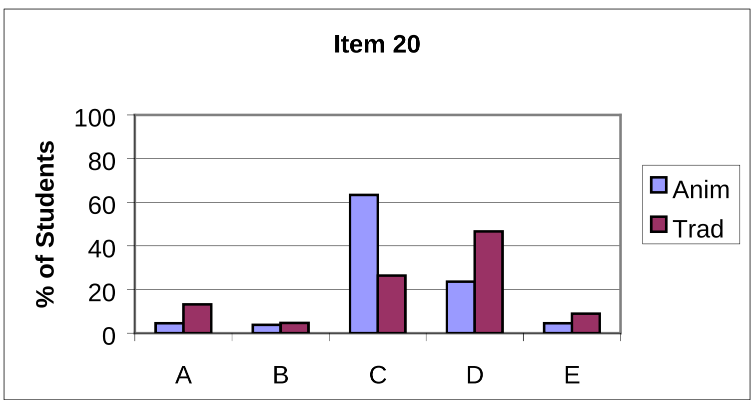


Figure 4.20 - Distribution of student answers for Item 20

Correct response = D

Choices with significant differences between animation and traditional = A, C, D

Table 4.20 - Summary of Statistical Data for Item 20

Item Twenty	Traditional	Animation
Correlation with English ACT	0.19	0.17
Correlation with Math ACT	0.38	0.22
% Correct	47	24
Index of discrimination	0.68	0.46

Points of Interest from the Interview Data

- Students may not know how to interpret the diagram in the traditional version.

Just as with Item 19, students who see the traditional version may have trouble understanding what the diagram represents. All three of the traditional students who answered incorrectly had difficulties understanding the diagram. For example Charlie stated,

> "For me it's not enough information because I've never been presented with this type of graph before. So I don't know what I'm looking at." (Charlie, Traditional)

I later asked Charlie to view the animated form of the question. He changed his answer to C (also incorrect) and stated that "the animation better depicted it." Janice also had trouble understanding the diagram in the traditional version. When I asked her to look at the animation she changed her answer from "Not enough information is given to answer the question" to the popular incorrect response C. She also commented that the animation helped.

> "Because I'm actually seeing it in motion. And it's easier to look at and actually be able to visualize when you see it happening and you see, oh well it's obvious that the green one is ahead of the blue one the whole time. Where as with just the little blocks it's a little, I don't know, I got a little confused. I wasn't sure like which one was ahead when." (Janice, Animation)

It is clear from the interview results for Items 19 and 20 that the strobe diagram is not a valid way to assess students' concepts of velocity and acceleration. Too many students are not able to express their ideas because they do not understand the question. For many students this question and Item 19 have more to do with their ability to read the diagram than their physics concepts.

- Students may assume the blocks start from rest in the animated version.

The animated version of Item 20 also has a fundamental flaw. When the animation is activated the two blocks are already present on the screen. As the animation plays out the blocks move across the screen. There were at least two students who assumed that the starting velocity of the blocks was zero. For example, Maggie identified that both blocks had the same acceleration as they crossed the screen.

> "They both look like they are about the same distance apart…the blue one's…it's like constant acceleration and it looks like the same for green too, only the green's just going faster." (Maggie, Animation)

But she chose the green block has having a higher acceleration because

> "It goes from…zero to whatever faster than, like where that one started." (Maggie, Animation)

She assumed that the blocks started from rest and concentrated on the accelerations of the blocks from this "rest" position. Lucy also assumed that the blocks stated from rest. After reading the answer choice D, "… both accelerations are zero" she commented

> "No because it went from zero to thirty-something. There had to be some acceleration there." (Lucy, Animation)

Both Lucy and Maggie demonstrated an imperfect understanding by interchanging position, velocity and/or acceleration. However, I believe that having the blocks start on the screen caused them unnecessary difficulties. The question would have been a more accurate representation of their understanding had the blocks come in from off screen.

- Velocity and acceleration are easily confused.

Many students interchanged the concepts of velocity and acceleration. This is not a surprising finding as this is the misconception Item 20 was designed to bring out. Example comments from students who made this error follow.

> "If it has a higher velocity, it has a higher acceleration." (Lucy, Animation)

> "The marks are wider for the green, which would mean it's accelerating faster, it's covering further distance in the same amount of time." (Alexia, Animation)

When I later asked Alexia to look at the traditional version of Item 20 she changed her incorrect response to the correct response. When I asked her why she had changed her response she could not tell me. She concluded that she must have made a mistake and had what she termed a "brain spaz".

Did the Animation have an Effect?

I believe the animation did affect the answers students gave to Item 20. I saw two different effects. First, the students had difficulty understanding what the diagram in the traditional version depicted. The animation had the effect of moving these students from an answer which was a guess to one which represented their understanding. The animation did not necessarily help them answer correctly, it just eliminated answers due to misunderstandings.

Second, the animation appears to steer students away from the correct answer to the incorrect answer C. When students see the animation, they are more likely to focus on the fact that one block is faster than the other when answering the question. When they see the traditional version (and interpret it correctly) then they are less likely to be obstructed by the faster velocity of one block and instead focus on their accelerations. It is not clear why the animation has this effect.

There were six students who saw the animated version of Item 20 and answered incorrectly. They all gave C as their answer choice. After completing the test, I asked four of these students to review the opposite form of the item. I asked Maggie and Beverly to review all 30 items in the opposite form. They both kept their answer to Item 20 without much comment as they did with almost all of the later items. I believe they were too tired at this point to seriously consider the alternative items I was presenting them, which was the reason I did not continue asking all students to review all items. Therefore, a low value should be placed on their lack of comment to the opposite form of this item.

The other two students were asked to review only a few items, with Item 20 being one of them. Both of these students took their review seriously and carefully looked at the traditional version of the item. Interestingly, both of these students changed their answer from the incorrect C to the correct answer D. Alexia was discussed above. She did not understand why she had originally given an incorrect answer. The other student to change his answer was Jason. When he saw the animation, Jason had correctly identified the blue block as having zero acceleration but had assumed that the green block did have an acceleration. When he saw the traditional version, he realized his mistake.

> "It looks like they have, it's got the same acceleration flat out. It's going to be equal this time. It looks like they're pretty much evenly spaced out." (Jason, Traditional)

I asked him why his answer was different, he replied

> "I think I saw it differently is because of the picture…This is the one instance where the picture made a difference. Because last time I think I was looking at it graphically as up here in a thing. And I think just looking at it differently, I was looking at the way the blocks were up here. But now that I actually see the actual numbers, the actual blocks, that I see it more spread out. That's probably why there is a difference." (Jason, Traditional)

Neither of these students were able to clearly articulate why they answered correctly when they saw the traditional version when they had answered the animated version incorrectly. Although it does appear that the animation had an effect, it is not clear why. I do not have enough data to support an explanative theory but I feel the answer probably lies with the sensitivity of student ideas. The ideas that students have are easily altered. As students' ideas move from a strong misconception to a strong understanding, they can pass through an in-between state were neither the misconception nor the correct understanding is strong in their minds. Students in this state may easily change the answers they give to conceptual questions with only slight differences in the way the question was asked. For example, consider Alexia's response to Item 20. She first answered the animated version of the question.

> "The marks are wider for the green, which would mean it's accelerating faster, it's covering further distance in the same amount of time." (Alexia, Animation)

She clearly has confused velocity and acceleration as she reasoned through her response. Later, when she saw the traditional version of Item 20 she stated,

> "The acceleration is fairly constant on both of them, or the speed is at least. So, well yes the speeds are constant and that means it's not accelerating." (Alexia, Traditional)

So after demonstrating the common misconception, she is able to turn around and demonstrate the proper understanding.

I believe that students with a strong understanding of the relationship between acceleration and velocity will answer either the animated or the traditional version correctly. Likewise, students with a strong misconception will consistently answer incorrectly. However, students with a weak understanding appear to be more likely to answer the traditional version correctly and the animated version incorrectly. I do not know why the traditional version is more likely to be answered correctly. Perhaps the answer lies within the psychological domain of perception.

Recommendation - Animation with Modifications

Both versions of Item 20 are flawed. The traditional version is not valid for the students who do not understand what the diagram represents. There are enough students with this difficulty that this flaw must be taken seriously, as it will have an overall effect.

The animated version has two problems. First, there is some confusion about the starting velocity of the blocks. But that can be eliminated by having the blocks enter from off screen as was discussed above.

Secondly, the animation seems to affect some students in such a way that they may answer incorrectly when they would have answered the traditional version correctly and been able to give a correct reason. Although it is not clear why this effect occurs, I believe it is due to the students' weak understanding as discussed above.

The inability of some students to understand the context of the traditional question is a serious shortcoming of the traditional version. Although there are some concerns associated with the animation, I believe it is superior for assessment because it better reflects student understanding. It should also be noted that both versions of this question have very high indices of discrimination, making it an excellent question of distinguishing between students.

4.3.21 Item Twenty-One

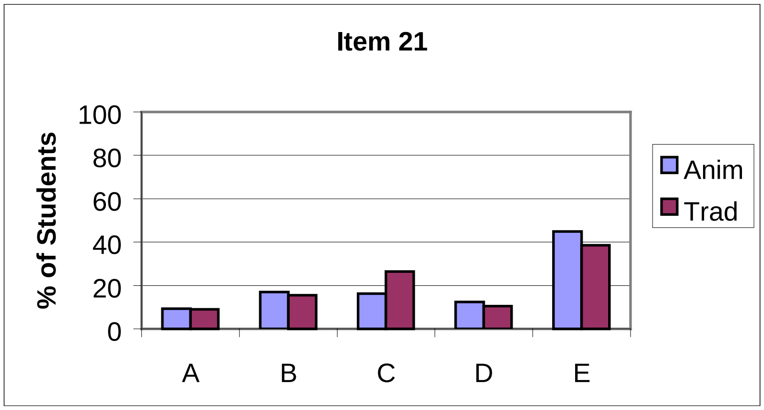


Figure 4.21 - Distribution of student answers for Item 21

Correct response = E

Choices with significant differences between animation and traditional = None

Table 4.21 - Summary of Statistical Data for Item 21

Item Twenty-One	Traditional	Animation
Correlation with English ACT	0.05	-0.11
Correlation with Math ACT	0.15	0.01
% Correct	39	45
Index of discrimination	0.40	0.17

Points of Interest from the Interview Data

- Item 21 does bring out student misconceptions.

There were several common misconceptions among the interview students.

1. Constant force (thrust) equals constant velocity (effect).

For example, Tanner (Traditional) chose answer C and stated, "Constant thrust will probably give it a constant velocity." Jason saw the animated version, also chose answer C and stated, "it's going to be simple vector addition once again."

2. Force of thrust must overcome the force of drifting.

This misconception is particularly interesting because it was demonstrated by students who answered the question correctly. The following statements were made by students who chose answer E.

> "When the force begins it's going to have to overcome the force of which it's going sideways, or whatever's keeping it there." (Charlie, Traditional)

> "It's going to gradually overcome the force of, just the direction that the rocket was going." (Alexia, Animation)

3. The force of the thrust slowly dissipates.

Georgia chose A as her answer. The reason she gave was,

> "…all the sudden the thruster comes on, and then it would gradually slow down as time went on…as time went on the power from the thrust would gradually decrease." (Georgia, Animation)

Did the Animation have an Effect?

In the large scale, there was no animation effect. With the exception of the index of discrimination, the statistical data shows no effect. There was only one interview student who was affected by the animation. Tanner answered Item 21 incorrectly in its traditional form. When later asked to look at the animated version he acknowledged he had made a mistake and changed his answer to the correct one. He said

"I was thinking of this as a position time, but I guess that really doesn't have to do anything with that…you think about a rocket taking off…it has to build up speed with the constant force." (Tanner, Animation)

Tanner felt that the animation had helped him.

Although I believe the animation did affect some students, it appears that the number of students affected is too small to detect on the large scale.

There was a fairly large difference in the index of discrimination with the traditional question being a better discriminator. However, I saw nothing in the interviews to explain this effect. It could be due to normal statistical fluctuations.

Recommendation - Traditional

Although Tanner's encounter with Item 21 does indicate that the animated version of this question may be superior, the evidence is weak. It is likely that the animation may help some students to understand the question statement but that the number of students affected by the animation is too small to be noticeable. Therefore, since the traditional version is simpler and easier to administer, it is recommended.

4.3.22 Item Twenty-Two

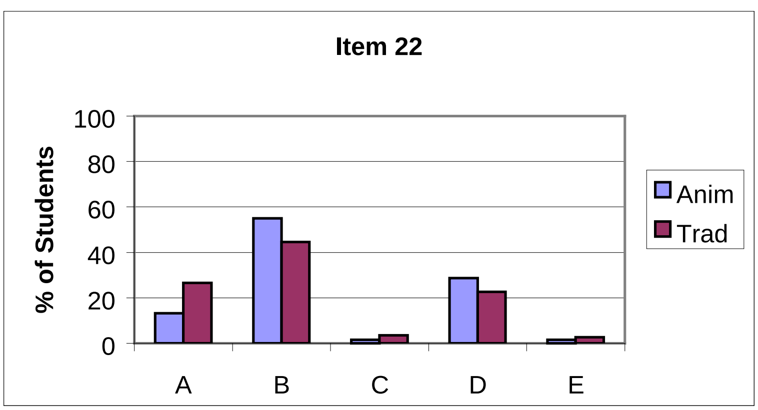


Figure 4.22 - Distribution of student answers for Item 22

Correct response = B

Choices with significant differences between animation and traditional = A

Table 4.22 - Summary of Statistical Data for Item 22

Item Twenty-Two	Traditional	Animation
Correlation with English ACT	0.19	0.01
Correlation with Math ACT	0.06	0.16
% Correct	45	55
Index of discrimination	0.33	0.42

Points of Interest from the Interview Data

- Students' answers are reflective of their understanding.

For the most part, the interview students who answered Item 22 correctly were able to provide sound explanations for their choices. For example,

> "If it's accelerating, which if there's constant forces on it then it is accelerating, then it's continually increasing." (Lucy, Animation)

> "If it's accelerating…which I guess it is since it's firing its rocket, then its speed is going to be continuously increasing." (Peter, Traditional)

Those who answered the question incorrectly had misconceptions. For example,

> "Because it's constant to me in space…there is no gravity in space…you see movies where [?] it's flung out into space and it will continue, I mean it will go on forever. Or they say that it will, at space camp." (Barb, Traditional)

> "There's no acceleration talked about, meaning what's good for one second is going to be good for the other there." (Jason, Animation)

Did the Animation have an Effect?

There is no indication from the interview data that the animation had an effect. The statistical data does indicate that the animation may have affected the proportion of students giving response A in the two groups. However, based on a complete lack of animation effect seen in the interviews, and due to the lack of a reasonable theory to support an animation effect for choice A, I believe the statistical difference seen was probably a false positive.

The lack of an animation effect for this question is actually very interesting. The question asks about the speed of the rocket. I had expected that students would try to determine the speed by simply watching the animation. Curiously, no student based their response on the animation. Even the students who were clearly unsure of their answers did not go back to the animation and attempt to retrieve the information.

Recommendation - Traditional

Since the animation does not appear to have any effect on students' answers to this question, the traditional version is recommended for simplicity.

4.3.23 Item Twenty-Three

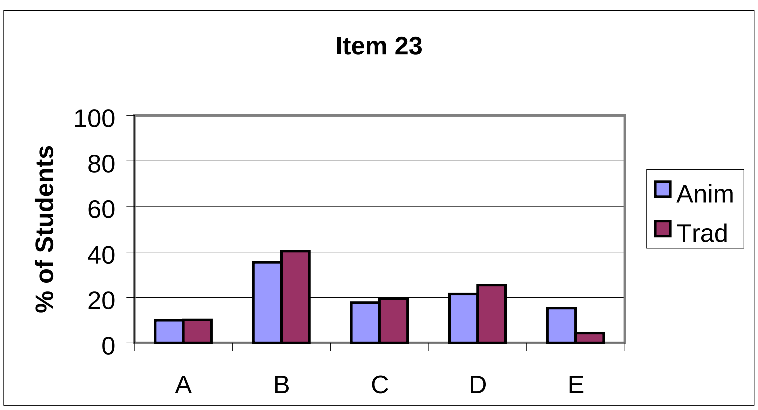


Figure 4.23 - Distribution of student answers for Item 23

Correct response = B

Choices with significant differences between animation and traditional = E

Table 4.23 - Summary of Statistical Data for Item 23

Item Twenty-Three	Traditional	Animation
Correlation with English ACT	0.12	-0.07
Correlation with Math ACT	0.20	0.03
% Correct	40	35
Index of discrimination	0.56	0.37

Points of Interest from the Interview Data

- Those who answered correctly generally did so for the correct reason.

For the most part, the students who answered correctly were able to demonstrate a correct understanding. For example,

> "Well since there is no forces I know that its path would be a straight line…I also know that it already had a [horizontal] speed and the thrusters gave it a [vertical] speed so my answer would be B." (Jake, Traditional)

> "I think it's going to just keep going the way it was going." (Maggie, Animation)

- False negatives are possible.

If a student answered Item 21 incorrectly they may answer Item 23 incorrectly even though they have a correct understanding. For example, Georgia gave the following reason for her choice to Item 23.

> "I was just thinking that since that was where it was headed when it was turned off, that it would just keep on going that way." (Georgia, Animation)

Her statement demonstrates that she understood that the rocket would continue along its path. But she had incorrectly answered A for Item 21. Since her answer for Item 23 was based on her choice for Item 21, her answer for Item 23 was wrong even though her reasoning was correct. It is not clear how widespread this problem is. It only occurred for one of my interview students.

- Students may assume gravity is part of the question.

Although the question statement places the rocket in space without outside forces, several interview students incorrectly accounted for the force of gravity. For example, Alexia chose D as her answer and stated,

> "When the thrusters are turned off it's going to slowly change course to be pushed back into orbit." (Alexia, Animation)

Tanner also chose D. His reason was similar to Alexia's.

> "I think it's D because it looks like force will start bringing it back down, or gravity." (Tanner, Traditional)

Did the Animation have an Effect?

There is no indication from the interview data that the animation had any effect. The statistical data does indicate that the animation students are more likely to chose E as an answer than the traditional students. Since this difference is completely unsupported by the interview data I believe that the statistical difference seen was due to chance and not a real effect.

Recommendation - Traditional

Since the animation does not appear to have any effect on students' answers to this question, the traditional version is recommended for simplicity.

4.3.24 Item Twenty-Four

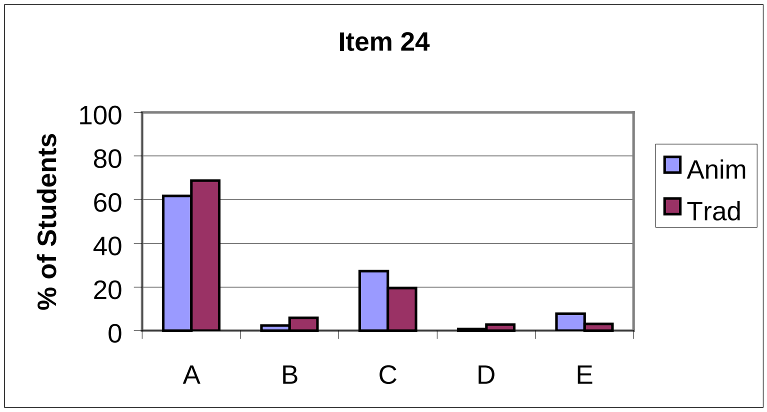


Figure 4.24 - Distribution of student answers for Item 24

Correct response = A

Choices with significant differences between animation and traditional = None

Table 4.24 - Summary of Statistical Data for Item 24

Item Twenty-Four	Traditional	Animation
Correlation with English ACT	0.06	-0.02
Correlation with Math ACT	0.05	0.15
% Correct	69	62
Index of discrimination	0.39	0.47

Points of Interest from the Interview Data

- Students who answer correctly do so for the right reason.

The students who answered Item 24 correctly all demonstrated an understanding of the correct rationale. For example,

> "The speed will not change since there is no outside forces. So it would be constant." (Jake, Traditional)

> "No force is acting on it, or friction or resistance." (Nathan, Animation)

- Students answering incorrectly have misconceptions.

The students who answered Item 24 incorrectly all demonstrated a weak, and often illogical understanding of the physical universe. For example,

> "I'll say E just because it's closest to what I'm thinking might happen, but I, I don't think it would decrease till where it would stop. It would just decrease to where it would still be constant, or where it was constant at first, the initial, the initial velocity." (Janice, Traditional)

> "It's going to have a little bit of thrust, but then afterward that thrust is going to be overcome by, I guess the force that had it in orbit, the force that had it drifting, the force going from A to B. So in that would change its direction and probably decrease its speed." (Alexia, Animation)

The most common error in thinking was that an object without a force will decrease its speed. This misconception is not surprising since real objects, in a world of friction, do decrease their speed when no longer pushed.

Did the Animation have an Effect?

There is no indication from either the statistical or interview data that the animation had any effect.

Recommendation - Traditional

Since the animation does not appear to have any effect on students' answers to this question, the traditional version is recommended for simplicity.

4.3.25 Item Twenty-Five

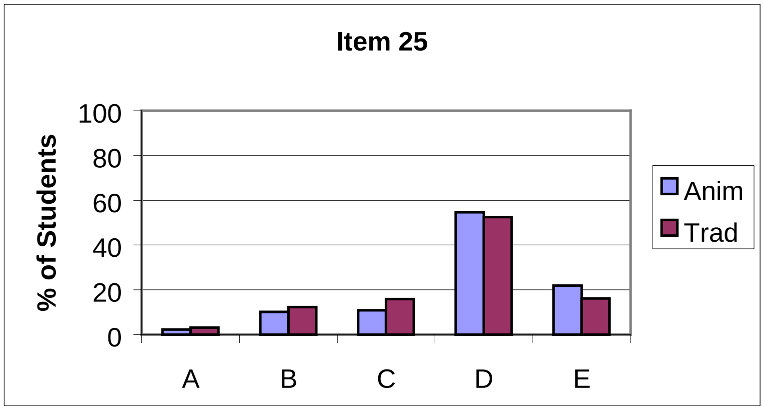


Figure 4.25 - Distribution of student answers for Item 25

Correct response = C

Choices with significant differences between animation and traditional = None

Table 4.25 - Summary of Statistical Data for Item 25

Item Twenty-Five	Traditional	Animation
Correlation with English ACT	0.04	-0.04
Correlation with Math ACT	0.04	0.22
% Correct	16	11
Index of discrimination	0.15	0.23

Points of Interest from the Interview Data

- The correct answer is reflective of the correct reasoning.

Although only two interview students answered Item 25 correctly, they both demonstrated a correct understanding. For example,

> "It's not accelerating so the sum of the forces would be zero." (Nathan, Traditional)

- Many students believe that there is a net force in the direction of motion.

Most of the interview students who answered incorrectly indicated that they believed the applied force would be greater than an resistive forces because the box was in motion. For example,

> "The force she's applying has to be greater than the force pushing back by the box because it, they are moving." (Jake, Animation)

This misconception is not surprising and has been seen in several other questions.

- Some students who saw the traditional version of the question considered the forces to get the box moving instead of the forces to keep it moving.

Interestingly, two of the students who saw the traditional version of Item 25 based their answer on the forces required to get the box moving rather than the forces required to keep it moving. For example, Alexia gave E as her answer and stated,

> "In order for the woman to get the box moving, she's got to exert more force, more forces than the…total force that resists its motion." (Alexia, Traditional)

Jason also answered E. He considered the resistive force to be static friction.

> "The constant horizontal force applied by the woman is greater than the total force that resists the motion of the box. So, I'm going to assume that's static friction." (Jason, Traditional)

Both Jason and Alexia answered E which includes the weight of the box as a resistive force. So their assumption that the question was asking about getting the box moving was not their only incorrect idea. However, it is not clear that they understood the basic question statement which means that their answers are not a valid measure of their understanding. Unfortunately, I did not ask either student to look at the animated version of Item 25. It would have been interesting to see if they recognized their error.

Did the Animation have an Effect?

There is no indication from the statistical data that the animation had any effect. However, based on the interviews, I believe the animation can help to clarify the question statement. As discussed above, two of the interview students who saw the traditional version may not have understood the exact nature of the question. None of the students who saw the animated version had this difficulty.

Recommendation - Traditional with Modifications

Although the animation may increase validity by increasing students' understanding of the problem statement, the traditional version could be modified to have the same effect. For example, the question could be changed to "…while the box is moving at a constant speed, the constant horizontal force applied by the women…". Hopefully this would eliminated misunderstandings.

The index of discrimination is very low but that is probably because the proportion of correct answers is low. Because this discrimination is based on a pre-test, and because there is no evidence of random guessing, the low discrimination is not a major cause for concern.

Since the traditional version can be modified to accommodate the benefits of the animated version, and because it is simpler, it is recommended.

4.3.26 Item Twenty-Six

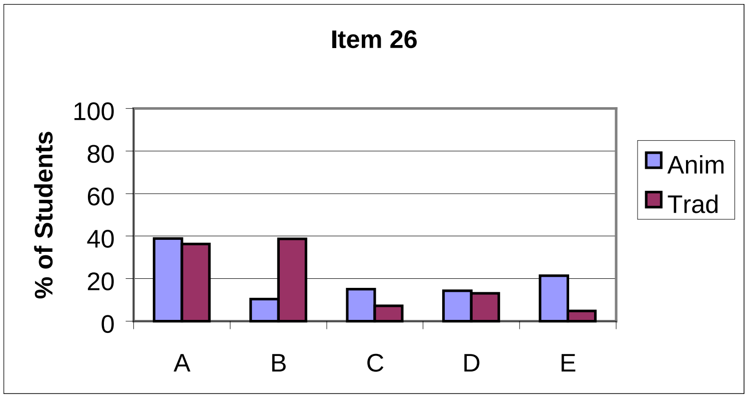


Figure 4.26 - Distribution of student answers for Item 26

Correct response = E

Choices with significant differences between animation and traditional = B, E

Table 4.26 - Summary of Statistical Data for Item 26

Item Twenty-Six	Traditional	Animation
Correlation with English ACT	0.03	-0.07
Correlation with Math ACT	0.03	0.07
% Correct	5	21
Index of discrimination	0.07	0.06

Points of Interest from the Interview Data

- A correct answer reflects a correct understanding.

Only 3 of the interview students answered Item 26 correctly. All of them demonstrated a correct understanding of the motion of the box. For example, even before viewing the correct animation Charlie stated,

> "It's going to be moving at a constant until it reaches point two then it's going to be steadily increasing." (Charlie, Animation)

He knew the correct motion of the box and just needed to find it among the animations. Nathan also understood why the box would increase its speed.

> "With a continually increasing speed because that would unbalance the sum of the forces and you would have a positive force creating acceleration." (Nathan, Traditional)

- Students who answered incorrectly had misconceptions.

All of the students who answered Item 26 incorrectly exhibited either an incorrect or a weak understanding of the effect of an added force on the box. They often assumed that the speed of the box would be constant after the second force would apply. This assumption would result in either A or B as an answer. It is not clear why they thought the speed would be constant after the second force was applied. The students simply stated it as fact without justification. For example,

> "[It] probably will keep the same constant velocity after time two seconds…[Animation E] looks like it accelerates. I don't think it accelerates except at that instance." (Tanner, Animation, Answered A)

Maggie's response was short. Her entire justification for her answer choice was,

> "I guess it would double it." (Maggie, Traditional, Answered A)

Did the Animation have an Effect?

An animation effect does not clearly stand out from the interviews. However, the statistical data indicated that the traditional group was more likely to give B as a response and the animated group was more likely to give the correct response E. With this in mind, a closer inspection of the interview data does reveal a possible animation effect.

For choice A, the animated version shows the box instantly jumping to twice the velocity at the point the additional force is applied. For choice B, the box also makes an instantaneous jump but only to a velocity that is 1.5 times the initial velocity. Some of the students who saw the animations could not distinguish this slightly higher velocity in Animation B from the initial velocity. In other words, they thought that Animation B depicted the box continuing at the same velocity after the additional force as its original velocity. For example, after viewing Animation B Peter stated,

> "I'm not sure whether it's going to go at constant speed after she doubles it…I don't, that seems to be the same rate at the end. I don't believe it's going to be the same rate." (Peter, Animation)

Peter chose D, which states that the speed of the box is constant after an increase in speed. Therefore, he does believe the speed should be constant at the end. In the above quote, he seems to be implying that the speed should not remain constant throughout. Tanner also seems to have run into the same problem. After viewing Animation B he stated,

> "B looks like it remains the same velocity or maybe it speeds up a little bit." (Tanner, Animation)

I later asked Tanner to look at the traditional form of Item 26. He changed his answer from A to B. His reason for the change was,

> "I guess I wasn't looking at the animation too closely because I didn't really recognize that the speed increases. I thought that it just kept at a constant speed once the second force was applied to it." (Tanner, Traditional)

Jason was also affected by the misperception of the velocity of the box in Animation B. Jason originally saw the traditional version of the question and answered D. I later asked him to look at the animated version. He changed his answer to C. But what was interesting was his reason for not choosing Animation B.

> "It looks like that's going to have the same speed as before…It looks like that one's just going to have a constant velocity straight through and that's not going to work." (Jason, Animation)

There were two other students who realized the box sped up in Animation B but thought it did not speed up enough to be correct. For example, Janice answered A and eliminated B because,

> "I think it would speed up a little more than that if you doubled the force." (Janice, Animation)

Interestingly, Janice changed her answer to B when I later asked her to look at the traditional version. Her reason for giving B on the traditional version was,

> "It would increase the speed but the force would just stay double. I don't think it would completely double the speed." (Janice, Traditional)

Alexia was opposite to Janice. She originally saw the traditional version of Item 26 and gave B as her answer because,

> "I don't think that there's a direct correlation to doubling the force and doubling the speed. If she's pushing more force then obviously she's going to increase the speed…but I don't know if it would necessarily be twice as great." (Alexia, Traditional)

I later asked her to look at the animated version. She changed her answer from B to E when she saw the animations. She discarded her previous answer B because,

> "I think the force, or the speed, is going to definitely increase more than Animation B, C, or D." (Alexia, Animation)

Alexia stated that she decided on Animation E because "it just looks more plausible." She could not justify choice E but felt it to be correct.

From the interviews it is apparent that the animation can deter students from choice B because the difference in the velocities before and after the push is small. Some students do not even detect the difference and others notice it but do not feel it is large enough. However, when these students are presented with the statement, "with a constant speed…but not necessarily twice as great" they are inclined to agree.

Besides Alexia's change to answer E on the animated version because E looked better, I did not see any reason why the animated group would be more likely to chose E. The students, on both versions, who originally gave E as an answer had a solid understanding of the concepts and did not seem to be affected by the presence or lack of an animation. It is possible that the significant difference in the proportion of students answering E on the two versions is a only a random statistical fluctuation.

Recommendation - Traditional

Very few students answered Item 26 correctly which may be the cause of its low index of discrimination. If this question is used in a post-test the results may improve. Otherwise, it appears to be a good question. There is almost no correlation with either math or verbal skills and the answers students give do seem to be reflective of their understanding.

Although the animation does appear to affect the proportion of students giving B as an answer, the effect is not clearly superior or hindering to assessment. In other words, the answers students give are generally equally reflective of their understanding in either version. Therefore, the traditional version of Item 26 is recommended for simplicity.

4.3.27 Item Twenty-Seven

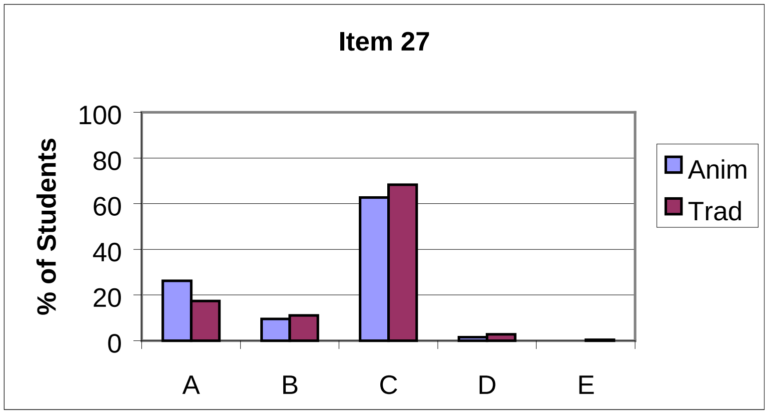


Figure 4.27 - Distribution of student answers for Item 27

Correct response = C

Choices with significant differences between animation and traditional = None

Table 4.27 - Summary of Statistical Data for Item 27

Item Twenty-Seven	Traditional	Animation
Correlation with English ACT	0.09	-0.08
Correlation with Math ACT	0.18	0.13
% Correct	68	63
Index of discrimination	0.33	0.14

Points of Interest from the Interview Data

- Students who answer correctly appear to have a correct understanding.

The majority of interview students answered this question correctly. They seem to have a correct understanding. For example,

> "C would look like something I would expect it to under friction condition." (Jake, Animation)

> "Its going to keep going but slow down…it's not going to keep going at the same velocity, because friction, there's a force resisting." (Tanner, Animation)

> "It would immediately start because there's the friction, the friction force that was against what she was pushing and that's immediately going to start slowing if there's not another force working on it." (Lucy, Traditional)

- There is a common misconception that force equals motion.

Only three of the interview students answered Item 27 incorrectly. They all gave A as their answer. All three students had the misconception that a force was required for motion and that there could be no motion without a force. For example,

> "It would not keep going without the force. She would have to exert extra force to make it go by itself, like give it a little push." (Janice, Animation)

> "If she stops applying that force there's going to be no momentum or anything carrying the box, because its weight and the forces of friction will stop it right away." (Alexia, Traditional)

Did the Animation have an Effect?

There is no indication for either the interviews or the statistical data that the animation had any effect.

Recommendation - Traditional

Item 27 appears to be a good question. The proportion of students answering correctly is neither high nor low and the index of discrimination is good. The answers students give are reflective of their understanding. Since there does not appear to be any animation effect the traditional version is recommended for simplicity.

4.3.28 Item Twenty-Eight

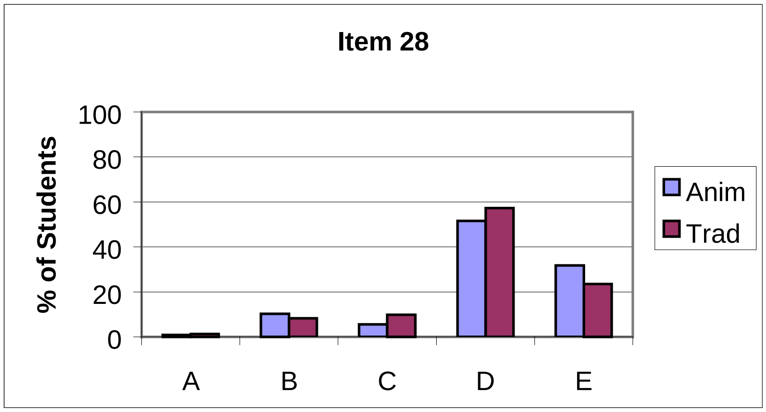


Figure 4.28 - Distribution of student answers for Item 28

Correct response = E

Choices with significant differences between animation and traditional = None

Table 4.28 - Summary of Statistical Data for Item 28

Item Twenty-Eight	Traditional	Animation
Correlation with English ACT	0.04	-0.22
Correlation with Math ACT	0.06	0.01
% Correct	23	32
Index of discrimination	0.34	0.51

Points of Interest from the Interview Data

- There are common misconceptions.

There were two misconceptions common among the interview students. First, many students believe the force of student "b" is either zero or less than the force of student "a" because student "b" is not active. They associated forces with active agents. For example,

> "I guess because "b" is just resistance, it's just something that's being pushed against. So he's exerting a resistance force, but he's not putting any action into it." (Alexia, Animation)

> "Student "b" does not exert any force on "a"…student "b" was just sitting there." (Beverly, Animation)

The second common misconception was that the more massive student would exert more force. It just so happened that student "a" was both more massive and the active agent so choice D would reflect both of these misconceptions.

- The question statement may confuse students.

In the question statement there is the phrase, "while the students are still touching one another." Two interview students who saw the traditional version interpreted this to mean that the students did not separate after the push. For example, Item 28 was one of the four questions Janice answered correctly. The reason she gave for the forces being equal was,

> "Both of them are exerting a force but it has equal amounts so neither of them are moving." (Janice, Traditional)

I later asked her to look at the animated version of Item 28. She changed her answer to D, which was more reflective of her understanding. She recognized that she had originally misinterpreted the question.

In contrast, Item 28 was one of only nine questions Charlie answered incorrectly. He stated that there were no forces between the two students because,

> "…after the push they're still together, moving…the only way I can see that feasible is if they didn't produce any force on each other." (Charlie, Traditional)

When I later asked Charlie to look at the animated version he changed his answer to the correct answer E "because the animation clarified it".

For these students the animated version was a better reflection of their actual understanding.

Did the Animation have an Effect?

Although there were no statically significant differences in the proportion of students giving particular answers on the two versions, I believe that the animation did have an effect. First, as discussed above, the animation clarified the question statement for some students. Secondly, some students may have looked at the animation to determine who was moving off faster as a method of answering the question. One of the interview students attempted to use velocity information and complained that he was not able to.

> "Since the people are kind of big compared to the size it wasn't easy to see, to tell who left the impact, who left moving faster. And it's off center too which makes it more difficult to tell who left at the greater impact." (Justin, Animation)

In the interviews, the most noticeable effect of the animation was to eliminate misunderstanding of the question statement. The statistical data does support this notion. First, the correlation between verbal ability and performance is smaller for the animated version. In fact, the correlation is actually negative for the animated version. This indicates that the students with less verbal ability are helped by the animation. Secondly, the index of discrimination is higher for the animated version. If the animation clears up misunderstandings about the question statement then it is a more valid measure. And if it is a more valid measure then it would be expected to have a higher index of discrimination.

Recommendation - Animation

The animated version of Item 28 appears to be a more valid measure of student understanding, therefore it is recommended. It should be noted that the statistical measures of correlation, proportion correct, and discrimination all indicated that Item 28, in its animated form, is a good question for assessment.

It is possible to modify the traditional version of Item 28 to eliminate the misunderstanding. However, the animated version is not time consuming since it contains only one simple animation. So the traditional version is not exceptionally simpler if other animations are already being used. Also, any attempts to modify the question statement may lead to further misunderstandings.

4.3.29 Item Twenty-Nine

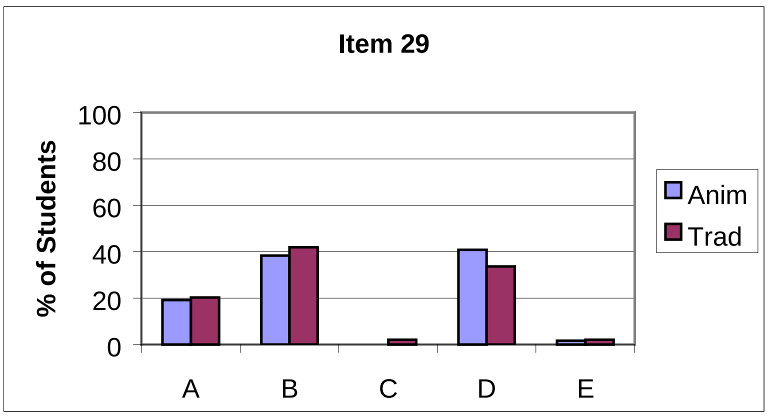


Figure 4.29 - Distribution of student answers for Item 29

Correct response = B

Choices with significant differences between animation and traditional = None

Table 4.29 - Summary of Statistical Data for Item 29

Item Twenty-Nine	Traditional	Animation
Correlation with English ACT	0.00	0.05
Correlation with Math ACT	-0.11	-0.02
% Correct	42	38
Index of discrimination	0.19	0.11

Points of Interest from the Interview Data

- The force of the air is interpreted differently by different students.

The "force of the air" does not mean the same thing to all students. Many of them are simply confused about what that force may be. For example,

> "A net downward force exerted by the air. We haven't gone through with that." (Jason, Traditional)

Others assume it must be asking about air resistance. For example,

> "I don't think that a downward force exerted by the air has any play in that because the object isn't going up. If it were, an empty office chair is flying through the air directly up, I might say that." (Alexia, Traditional)

> "The force of the air only works in air resistance." (Charlie, Animation)

And others attribute it to air pressure. For example,

> "Since air does have weight, it would also be having, it would be like stacking a pillar of air all the way to wherever it is on top of the chair." (Jake, Animation)

> "I suppose that the downward force by the air would be [?], that's like air pressure." (Peter, Animation)

Since students interpret the "force of the air" differently, the responses given do not have the same meaning for all students.

Did the Animation have an Effect?

There is not indication that the "animation" had any effect. This was not surprising as the animation is actually a static picture of an object all students are presumed to be able to visualize clearly in their heads.

Recommendation - Traditional with Modifications

The animation had no effect so there is no need to use it. However, I believe the question could be improved by clarifying what the force of the air is meant to be. Also since almost no one chose C or E as their answer, those choices could be used to expand the question. The following is what I suggest as the modified question.

An empty office chair is at rest on a floor. Consider the following forces:

1. A downward force of gravity.
2. A upward force exerted by the floor.
3. A net downward force due to air pressure.
4. A net upward force due to air pressure.

Which of the forces is(are) acting on the office chair?

(A) 1 only

(B) 1 and 2

(C) 1, 2, 3

(D) 1, 2, 4

(E) None of the forces, since the chair is at rest there are no forces acting on it.

If the above question is used it will be easy to distinguish between those students who think air pressure has an effect and those who do not. Also, it will demonstrate which direction students believe the air pressure will act.

4.3.30 Item Thirty

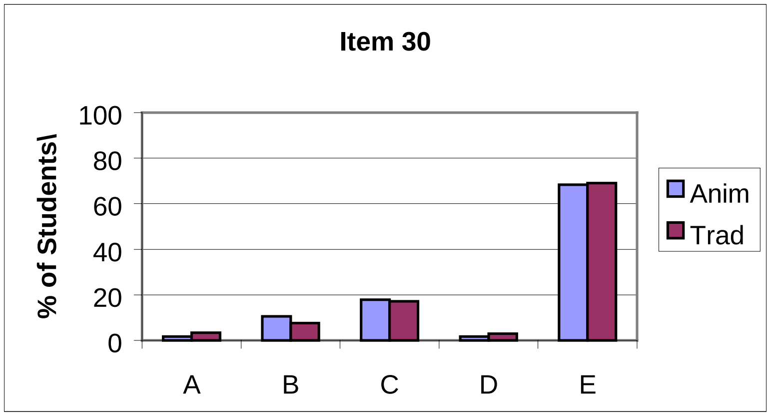


Figure 4.30 - Distribution of student answers for Item 30

Correct response = C

Choices with significant differences between animation and traditional = None

Table 4.30 - Summary of Statistical Data for Item 30

Item Thirty	Traditional	Animation
Correlation with English ACT	0.07	0.07
Correlation with Math ACT	0.17	0.20
% Correct	17	18
Index of discrimination	0.33	0.45

Points of Interest from the Interview Data

- There is a common misconception that the force of the hit is acting on the ball after it leaves the racquet.

Unsurprisingly, many students demonstrated that they believe the force of the hit was acting on the ball as it passed over the net. For example,

> "It says after it has left contact with the racket, still the force of the hit is acting on it." (Peter, Traditional)

> "Maybe 1, 2, and 3. I mean it is going through the air and it was hit." (Beverly, Animation)

Did the Animation have an Effect?

There is no indication that the animation had any effect.

Recommendation - Traditional with Modifications

Since the animation had no effect, the traditional version is recommended for simplicity. Item 30 appears to be a good question. The answers students gave were reflective of their thinking, the correlation is verbal skill is zero, the correlation with math skills is low and the index of discrimination is good.

There was one interview student who wanted to include a normal force but could not since it was not a choice. Since very few students gave A or D as an answer choice, I suggest modifying the question to include a fourth force, "a normal force". The answer choices could then be:

(A) 1, 2, 3 and 4

(B) 1 and 2

(C) 1 and 3

(D) 1, 3, and 4

(E) 1, 2, and 3

I believe this small change would fine tune this already acceptable question.

4.4 Summary of Findings from the Interviews

The interviews provided a great deal of detail about the way students interact with each question. But there were several common findings that can be generalized to the test as a whole, and possibly to conceptual assessment in general. In the following sections I discuss those findings that are most relevant to conceptual assessment in general.

4.4.1 Findings not Related to the Animations

Although it was not my purpose to investigate anything beyond the use of animations in assessment, the data I collected and the analysis performed have led to findings about general conceptual assessment. There were two findings which were particularly striking. Both have been briefly reported by the authors of the FCI, but they are important enough to be restated here.

- There is almost no random guessing on the FCI.

Multiple choice exams are often necessary when large numbers of students need to be assessed quickly and with consistency. However, one major drawback to the multiple choice exam is that it makes random guessing easy. On a test like the FCI with 5 answer choices, a student could score around 20% through random guessing alone. Considering that the average FCI score is only slightly higher than the score based on random guessing, there is the potential for great concern.

However, this concern is unjustified. During the interviews, I saw no instance of a completely random guess on any question. There were only a handful of times when a student seemed so unsure about an answer that I would describe his/her choice as a guess among two or three answer choices. Overwhelmingly, the students had definite reasons for each choice that they made. This adds a lot of strength to assessment with the FCI.

- Student views are inconsistent.

Students' views about our physical world are not black and white. They do not have one consistently wrong view and then suddenly move consistently to the correct view. Their incorrect views are based on illogical ideas and will fail them from time to time. This will cause them to abandon or modify their views in a way that can appear random to an expert. Even once students are exposed to the correct view, and begin to apply a correct understanding, they will still revert to their previously help beliefs on occasion. Many of the students I interviewed demonstrated turmoil over the belief they wanted to be true and the correct view that they were beginning to understand. This process of conceptual change is very fascinating.

This inconsistency of beliefs added a definite dimension of error to measurements made with the FCI. Students would answer two "identical" questions differently. By identical, I mean questions that asked about the same concept. In other words, questions that an expert would consider to be essentially the same. For example, on Item 5 (forces on a ball in a circular tube) Jason included the force in the direction of motion,

> "A force in the direction of motion. That's going to be the, I'm trying to think of the proper word for that, tangential acceleration." (Jason)

But on Item 11 (forces on a sliding hockey puck) he did not.

> "And gravity's the only thing working on it because there's no horizontal force in the direction of motion." (Jason)

He also disregarded the normal force on Item 11,

> "There's no upward force exerted by the surface because we were just assuming there is no friction." (Jason)

while he included it in Item 29 (forces on an office chair), "An upward force exerted by the floor, that's the normal force." (Jason)

Like Jason, the students did not always recognize their own inconsistencies. However, many students did. For example, Beverly did not include the normal force on Item 11 (forces on a hockey puck).

"…the upper force [pause] I don't think it's going to have anything to do with the upper force." (Beverly)

When she got to Item 29 she realized her error,

"I think [my high school physics teacher] was saying something about upward force, but I can't remember…If I say [the upward force is included] then it means I'm going to have to find the other question and change its answer." (Beverly)

Beverly was torn between her basic belief and what she had learned in class. This was not uncommon. Other excellent examples of this can be found in the discussion of Item 1.

These students were not stupid and were putting thought into answering the questions correctly. But they were struggling with their beliefs. Although they did not randomly guess, their conceptual discord did add an element of randomness to their answers.

4.4.2 Findings Related to the Animations

- Animation can reduce misunderstandings about the question statement.

The most significant result to come out of this project is probably that **animation can reduce the link between reading ability and question response**. The statistical analysis discussed in Chapter 3 indicated that performance on the animated questions was less correlated with verbal ability than performance on the traditional version of the question. The interviews confirmed that the animations reduce errors due to poor reading comprehension.

There were a number of instances where students clearly misinterpreted the question statement. Examples can be found in the previous discussions of Items 8 and 28. In these cases, the questions were clearly stated and the errors were due to the students' poor reading of the question.

Animation can also reduce question misunderstanding due to vague question statements. Item 4 provides an excellent example. As was discussed, this question left room for interpretation, making the question slightly different for each student. In the animated version, the situation of the question was clear. This makes the animation superior.

One might argue that the question could simply be rewritten for clarity, and in this case that is true. However, it is unusual to have a question for assessment that has been so thoroughly researched as those found on the FCI. Usually questions are written for assessment and then given to students without extensive study for understanding and validly. It is not possible for the classroom teacher to delve so deeply into each question he or she asks of the students. Therefore, in any case where the possibility of an unclear question statement exists the animation can be useful.

It is important to note that animation does not necessarily lead to improved performance. What it does is increase the validity of the assessment. So, a student with poor verbal skills will not necessarily be "helped" by the animation. The animation will only increase the performance of that student if he or she had a correct understanding that was not reflected due to poor reading skills. In some cases, the animation may actually decrease the performance of students with reading difficulties. This can be seen in the discussion of Janice's response to Item 28. She gave the correct answer to the traditional form of the question, even though her understanding was faulty, because she misread the question statement. When she saw the animation, she gave an incorrect answer that was true to her beliefs.

- Students respond to the animation on a more intuitive level than they respond to static questions.

As students answer an animated question, they must process information differently. For example, instead of being told an object moves with constant velocity, they must interpret constant velocity from its motion. This interpretation is not necessarily literal, meaning that the students can use the information about the velocity being constant without actually having to make it explicit in their minds.

One effect of this more intuitive interaction was seen in Item 1. As previously discussed, the results of Item 1 indicate that the animation may reduce memorized responses. It appears that the animation can be useful for taking students out of school mode and deeper into their own thinking.

Interestingly, students' intuition is often surprisingly incorrect. When I first wrote the animated questions and gave them to students, I was sure that EVERYONE would answer some of the questions correctly. For example, anything but the correct path on Item 14 (ball falling from an airplane) looks extremely awkward to me in the animated version. The incorrect paths did not seem quite as unreasonable in the traditional version. I was shocked when on this, and other questions, the students in the animated condition still gave incorrect answers.

The animated group was more likely to give the correct answer to Item 14 than the traditional. But, much to my surprise, 25% of the students still gave an incorrect answer. I believe this is an indication of a very deeply rooted misconception and an indication that the results are reflective of students' understanding and not an artifact of the way the question is asked. If a student has a very strong view, then he or she will be able to deny evidence to the contrary such as the obvious backward kick of the ball as it falls behind the plane in path A.

So although the animation does tap into basic intuition, it does not eliminate strong misconceptions. This is good news for assessment. It means that animations do not give the answer away as might be thought. In theory, this aspect of animation could be used as a honed assessment of beliefs since students with strong views will answer consistently in either version but students with transforming views will not.

- Perception is important.

An alternate explanation for the lower than expected performance on some of the animated questions lies in perception. This project was not designed to look at issues of perception. But it is clear that it will be important to do so if animations are used extensively for assessment purposes. A fairly complex issue of perception arose in Item 14 as is discussed there.

On a simpler level, it is apparent that students are not always able to determine important information by viewing the animation. For example, students were not always able to detect the velocity change in Item 26. In designing animations care must be taken to ensure that important quantities can be perceived.

It is also apparent that perceptual issues can make seemingly minor differences in presentation have large effects. For example, students on Item 20 were affected by the starting of blocks with non-zero velocity on-screen instead of off-screen. And, there is some evidence that the leaving of footprints or a trail can alter students' responses.

- With the exception of the extra time involved, the animations were rarely found to be destructive to assessment.

In terms of assessment, the animations were almost always as good or better than their traditional counterparts. They reduced misunderstandings and were more connected to students' intuition.

There were only two instances where the animation was inferior to the traditional questions. The first was the confusion caused by having Item 5 appear before Item 6 which was discussed earlier. This confusion could have easily been avoided and only existed because of the attempt to keep the animated version as close as possible to the traditional version. The other instance where the animation was cause for concern was in Item 26. In that question students could not always detect the change in velocity, which has also been previously discussed.

This result has promising implications for using animations in assessment. Overall, the animations appear to improve assessment.

The drawback is that developing the animations takes more time than the writing of a traditional question. Also, animated question are more difficult to give to a group of students because each student needs access to a computer. Unfortunately, computer access is still limited in most locations. Finally, many animated question take students longer to answer because the animation takes longer to play than the traditional question would take to read.

So although the use of animations for assessment does show promise, their value must be judged in relation to their ease of use.

---

[1] Patton, M. Q. (1990). Qualitative Evaluation and Research Methods (Second Edition). Newbury Park: Sage Publications.
[2] Strauss, A. & Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park: Sage Publications.

Chapter 5
Conclusion

There were three main questions which I intended to address with this project. Below is a summary of findings for each question.

5.1 Research Question 1

When static pictures and descriptions of motion are replaced with computer animations, how are students' responses to Force Concept Inventory questions affected?

- In general there was no effect. (Section 3.3.1)

This result is important because it indicates that animation should not be used just because it is possible. The tendency to use a technocentric approach should be avoided.

- Animation can either improve or decrease performance. (Section 3.3.1)

It might be thought that animation would only increase performance since it eliminates the cognitive step of interpreting words. However, this was not found to be the case. The connection between presentation of question with or without animation and performance is apparently not straightforward.

- An effect is most likely when the question involves motion and the animation is necessary to answer the question. (Section 3.3.2)

Although obvious in hindsight, this is an important result as it can serve as a guide when choosing questions or tutorials to be animated.

- In general, performance on animated questions is less linked with verbal ability than performance on traditional questions. (Section 3.4)

There are several implications of this very important result. First, it implies that animation may be particularly useful with populations with reading difficulties. For example, international students, very young students, students with learning disabilities, or students who simply have less verbal ability.

Secondly, animations could be used to increase the validity of questions since they tend to diminish the effect of a confounding variable. The animated version of a question is less likely to measure reading ability.

- Animation may reduce the gender gap. (Section 3.5)

There are indications that the performance of men and women will converge on animated questions. However, the evidence is not strong and this finding would be worthwhile to investigate further. As of this writing, the gender gap is well documented but not understood. Animations may provide a way to probe this area.

5.2 Research Question 2

If student responses are affected by the animation, what are the reasons behind these effects?

It turned out that there are many reasons an animation can effect the outcome of an assessment question. The effect of an animation can be complicated and no finding can explain all animation effects. The following are the most common reasons I saw for an animation effect.

- A traditional question can be misunderstood.

This was seen in Questions 14, 19, 20, 25, and 28. In each of these questions students did not always answer the question that was being asked because they did not understand the question. In each case, the animation cleared up the misunderstandings.

- A traditional question can be vague.

Question 4 demonstrates the best example of this. The statement was vague enough that students could interpret it in different ways. Although all the information an expert could consider necessary was in the question statement, the students often wanted to use information that was not included.

- Relevant quantities in an animation may not be noticeable.

This was only an issue in one question, Question 26. When questions are animated, care should be taken that the boundaries of perception are not approached.

- Students are more likely to respond based on their intuition and less on what they have learned in school.

Questions 1 and 2 provide the best examples of this effect. For both of these questions I believe the animation better assessed the students by decreasing memorized responses.

- Animation can help students visualize a situation.

There are times when a student will answer an animated question differently from a traditional question because the animation makes a particular choice look unnatural. For example, in Question 3 the animation students were not likely to state that an object in free-fall accelerates briefly then falls at a constant rate. When these students actually saw the motion described they recognized it as being incorrect. Presumably, they were unable to sufficiently visualize the motion in their heads to use as a tool for answering the question in its traditional format.

5.3 Research Question 3

What implications can be drawn about the value of traditional and animated versions of the Force Concept Inventory for conceptual assessment of physics concepts?

The implications for each question are discussed in detail in Chapter 4. However, a number of general implications were found.

- The value of animations for assessment is usually either an improvement or a tie with traditional counterparts.

In most cases, it was found that the animation either had no effect on the value of a particular question or it improved the question. This is an important result because it means that animations can generally be used without fear of a destructive influence. If the difficulty of question construction and ability to administer is not a concern, then animations can be used without too much worry of adverse effects.

- Animations are often more valid.

Because of problems with question misunderstanding and vagueness in the traditional questions, the animation often provided a more valid measure. This finding is very important as it provides a very strong argument in favor of using animations for assessment.

- Both the animated and the traditional FCI are reliable.

In terms of reliability, neither version was superior.

5.4 Summary of Important Results

The following is a list of those findings that I believe to be most important and/or interesting. Each of the findings is discussed in detail elsewhere in this dissertation.

- Animation is most likely to have an effect when the question involves motion and the animation is central to answering the question.
- Performance on animated questions is less linked to verbal ability.
- Animations can reduce question misunderstandings, making them a more valid measure.
- Animation is rarely a detriment to assessment.

5.5 Future Research

Although I have answered many interesting questions, and provided insight into numerous areas, even more questions have developed. Since almost nothing had previously been done in the area of animations for assessment, my goal with this project was to lay the ground work in this area. The basic building blocks have now been laid and the next step is to investigate more specific areas and findings from this project. The following are those questions which I feel would be most worthwhile to investigate.

- Are there gender differences in the way men and women respond to an animation?

As I have discussed, I found indications that animation may reduce the gender gap. If this result is accurate then it will be very important. There is an awareness that women do not perform as well as men on the FCI, even when general academic ability is accounted for. Understanding the interaction between animation and gender could add valuable insights into this area.

- Is the effect of animation the same for students of all abilities and backgrounds?

This project focused on a very specific group of students. I have reason to believe, from data that was not part of this dissertation, that the effect of animation may be sensitive to the background and abilities of the students. I believe it would be worthwhile to look at this matter further.

- What is the effect of the trail/footprints left by an object?

There are a number of indications that the record of an objects motion sometimes left on the screen could have affected the outcome of student responses. As discussed in Chapter 3, McCloskey and Kohl have suggested the existence of an object trail could diminish the effect of an animation by having students focus on the static diagram.

I have also seen evidence of this in my project. As I discussed in Chapter 4, there is the possibility that the trail, or lack of a trail had an effect on Question 6 and 7. Also, a number of the students used the footprints on Questions 19 and 20 as the basis for their answer instead of concentrating on the motion. I believe this question had the potential to give interesting and valuable results, especially for instructional software designers.

- To what extent does the FCI measure perception over cognitive understanding?

There is some evidence, particularly on Question 14 (ball from an airplane) that the answers students give have more to do with perception than with their actual understanding of the physical universe. Given that some of the FCI questions are also found in the psychology research literature, it is possible that there are issues surrounding these questions that should be investigated further.

- Could a hybrid FCI be developed using the results of this study that would be an improvement over the current version?

This is the logical next step of this project. Now that I have identified the good and the bad of both forms of each question, a hybrid FCI of the best of both versions should be created and tested. I believe that such a hybrid would be a superior assessment instrument.

- How can animations be used in assessment in unique ways?

I did not set out to develop the best conceptual instrument. My goal was to learn more about how animations could be used. In order to do this I intentionally used animated questions that had a traditional parallel. I do not believe I have made the best use of the technology available to us.

We currently use a paradigm of testing that is centered on the possibilities offered by paper. This method of question delivery and response has certainly shaped the type and nature of the questions we ask. In fact, paper assessment is so much a part of our education system that it is difficult to even imagine alternatives.

But we are no longer limited to paper based assessment and it is time we started to think outside the proverbial box. Computer animations open doors never possible before. For example, they can test students through interactivity, constantly adapting to the students' input. They can also ask questions in unique ways. For example, a computer animation could be used to ask students "Is this object speeding up?" Also, animations can mimic real situations by giving students more information than they need and by not highlighting necessary information by stating it explicitly.

The possibilities are immense and difficult to imagine. By breaking from our current mode of assessment, we could potentially make drastic improvements in our abilities to measure student understanding.

5.6 Concluding Statement

Based on my work, I believe that animation can improve assessment in many ways. The potential for improvement is great but the current state of investigation is small. I believe I have laid the basis for future research by illustrating areas where animation can be most valuable and by offering insights into the reasons behind the effects.

I believe this area of research is growing in importance as the availability of computers in classrooms increases. I also believe there are many interesting and worthwhile questions to be answered. The most interesting and worthwhile will probably come from the application of animation beyond a paper parallel. It will be exciting to watch this new paradigm of assessment emerge.

Appendix A – The Animated Force Concept Inventory as seen by students.

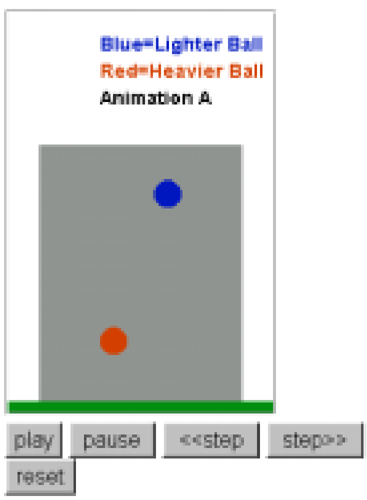
Force Concept Inventory

Please:
Do not skip any question.
Avoid guessing. Your answers should reflect what you personally think.
Enter the following information before you begin.

Your Name(Last, First):
Your Student ID:
Your School:
Your Instructor:

IMPORTANT!!!!!!! Before you view any animation you must first click on the TEXT "Start Animation" to start the animation. After you have clicked Start Animation the animation will run and can be viewed using the control buttons below the animation.

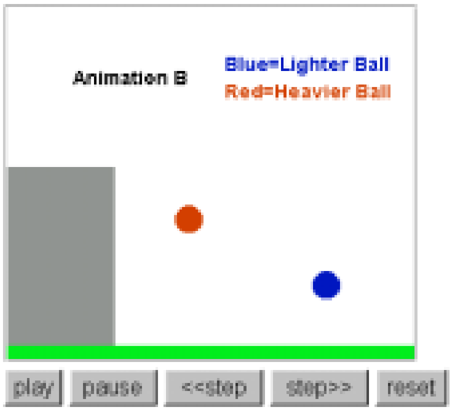
Question 1



Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. Which animation best represents how the balls would fall to the ground below?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

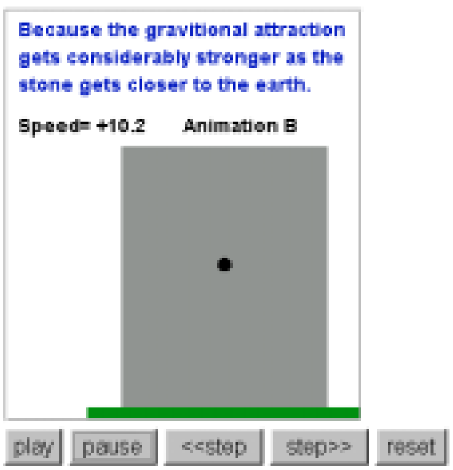
Question 2



The two metal balls of the previous problem roll off a horizontal table with the same speed. Which animation best represents the distance each ball will land from the table?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 3



Which animation best represents the way a stone dropped from the roof of a single story building falls to the surface of the earth?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 4

Start Animation

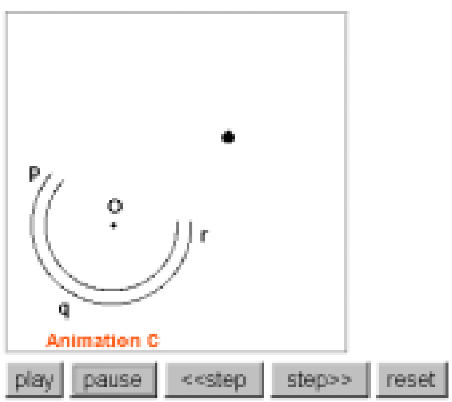
A large truck collides head-on with a small compact car as shown in the animation. During the collision:

- (A) the truck exerts a greater amount of force on the car than the car exerts on the truck.
- (B) the car exerts a greater amount of force on the truck than the truck exerts on the car.
- (C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.
- (D) the truck exerts a force on the car but the car does not exert a force on the truck.
- (E) the truck exerts the same amount of force on the car as the car exerts on the truck.

USE THE STATEMENT AND ANIMATION BELOW TO ANSWER THE NEXT TWO QUESTIONS (5 AND 6).

The accompanying animation shows a frictionless channel in the shape of a segment of circle with center at "O". The channel has been anchored to a frictionless horizontal table top. You are looking down at the table. Forces exerted by the air are negligible. A ball is shot at high speeds into the channel at "p" and exits at "r".

Question 5



Consider the following distinct forces:

1. A downward force of gravity.
2. A force exerted by the channel pointing from q to O.
3. A force in the direction of motion.
4. A force pointing from O to q.

Which of the above forces is(are) acting on the ball when it is within the frictionless channel at position "q"?

- (A) 1 only.
- (B) 1 and 2.
- (C) 1 and 3.
- (D) 1, 2, and 3.
- (E) 1, 3, and 4.

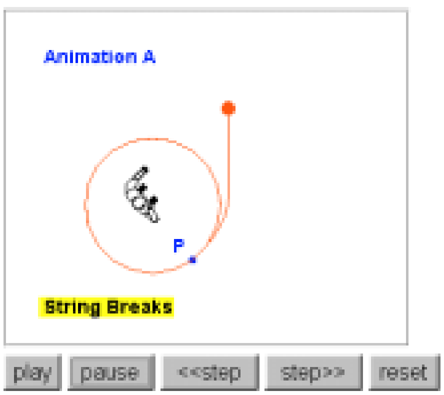
Start Animation A   Start Animation B   Start Animation C   Start Animation D   Start Animation E

Question 6

Which animation shows the path the ball would most closely follow after it exits the channel at "r" and moves across the frictionless table top?

- (A)
- (B)
- (C)
- (D)
- (E)

Question 7



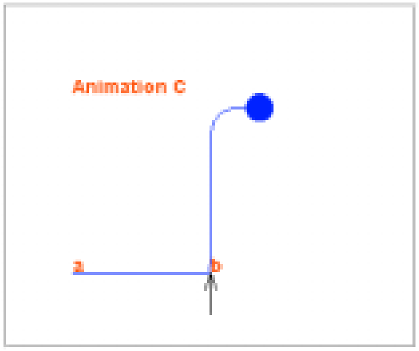
A heavy ball is attached to a string and swung in a circular path in a horizontal plane as shown in the animation. At the point P indicated in the animation, the string suddenly breaks near the ball. If these events are observed from directly above, which animation shows the path the ball would most closely follow after the string breaks?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

USE THE STATEMENT AND ANIMATION BELOW TO ANSWER THE NEXT FOUR QUESTIONS (8 through 11).

The animations depict a hockey puck sliding with constant speed v_0 in a straight line from point "a" to point "b" on a frictionless horizontal surface. Forces exerted by the air are negligible. You are looking down on the puck. When the puck reaches point "b", it receives a swift horizontal kick in the direction of the black arrow. Had the puck been at rest at point "b", then the kick would have sent the puck in horizontal motion with a speed v_k in the direction of the kick.

Question 8



play pause reset

Which animation shows the path the puck would most closely follow after receiving the "kick"?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 9

The speed of the puck just after it receives the kick is:

- (A) equal to the speed "v_o" it had before it received the kick.
- (B) equal to the speed "v_k" resulting from the kick and independent of the speed "v_o".
- (C) equal to the arithmetic sum of the speeds "v_o" and "v_k".
- (D) smaller than either of the speeds "v_o" or "v_k".
- (E) greater than either of the speeds "v_o" or "v_k", but less than the arithmetic sum of these two speeds.

Question 10

Along the frictionless path you have chosen in question 8, the speed of the puck after receiving the kick:

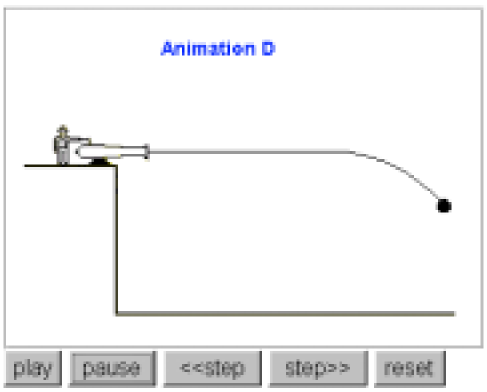
- (A) is constant.
- (B) continuously increases.
- (C) continuously decreases.
- (D) increases for a while and decreases thereafter.
- (E) is constant for a while and decreases thereafter.

Question 11

Along the frictionless path you have choosen in question 8, the main force(s) acting on the puck after receiving the kick is (are):

- (A) a downward force of gravity.
- (B) a downward force of gravity and a horizontal force in the direction of motion.
- (C) a downward force of gravity, an upward force exerted by the surface, and a horizontal force in the direction of motion.
- (D) a downward force of gravity and an upward force exerted by the surface.
- (E) none. (No forces act on the puck.)

Question 12



A ball is fired by a cannon from the top of a cliff as shown in the animations. Which of the animations best represents the path the cannon ball would most closely follow?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 13

A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that the forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):

Start Animation

(A) a downward force of gravity along with a steadily decreasing upward force.

(B) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth.

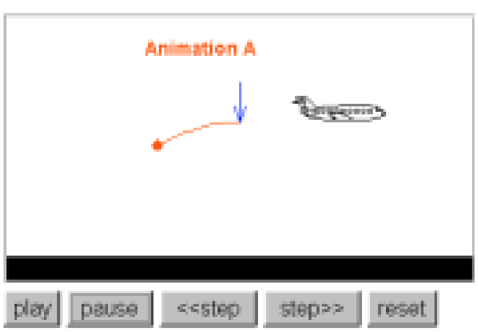
(C) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity.

(D) an almost constant downward force of gravity only.

(E) none of the above. The ball falls back to the ground because of its natural tendency to rest on the surface of the earth.

Question 14

A bowling ball accidently falls out of the cargo bay of an airliner as it flies along in a horizontal direction. As observed by a person standing on the ground and viewing the plane as shown in the animations, which animation shows the path the bowling ball would most closely follow after leaving the airplane? (note that the bowling ball falls out at the point indicated by the blue arrow)



Start Animation A

Start Animation B

Start Animation C

Start Animation D

Start Animation E

USE THE STATEMENT AND ANIMATION BELOW TO ANSWER THE NEXT TWO QUESTIONS (15 and 16).

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the animation.

Start Animation

Question 15

While the car, still pushing the truck, is speeding up to get to cruising speed;

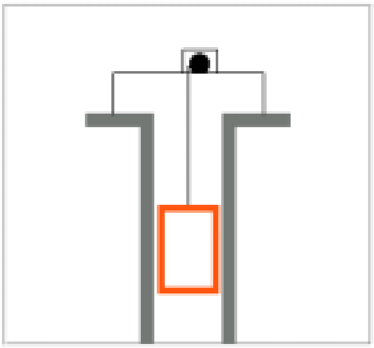
- (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
- (B) the amount of force with which the car pushs on the truck is smaller than that with which the truck pushs back on the car.
- (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
- (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck can not push back against the car. The truck is pushed forward simply because it is in the way of the car.
- (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

Question 16

After the person in the car, while pushing the truck, reaches the cruising speed at which he/she wishes to continue to travel at a constant speed;

- (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
- (B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushs back on the car.
- (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
- (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck can not push back against the car. The truck is pushed forward simply because it is in the way of the car.
- (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

Question 17



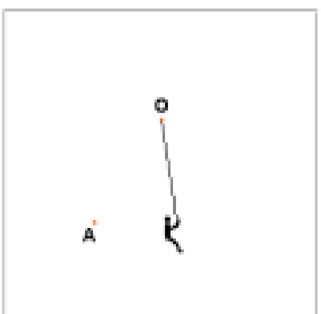
play | pause | <<step | step>> | reset

Start Animation

An elevator is being lifted up an elevator shaft at a constant speed by a steel cable as depicted in the animation. All frictional effects are negligible. While the elevator is rising, forces on the elevator are such that:

- (A) the upward force of the cable is greater than the downward force of gravity.
- (B) the upward force of the cable is equal to the downward force of gravity.
- (C) the upward force of the cable is is smaller than the downward force of gravity.
- (D) the upward force of the cable is greater than the sum of the downward force of gravity and the downward force due to the air.
- (E) none of the above (The elevator goes up because the cable is being shortened, not because an upward force is exerted on the elevator by the cable).

Question 18



play | pause | <<step | step>> | reset

The animation shows a boy swinging on a rope, starting at a point higher than A.
Start Animation

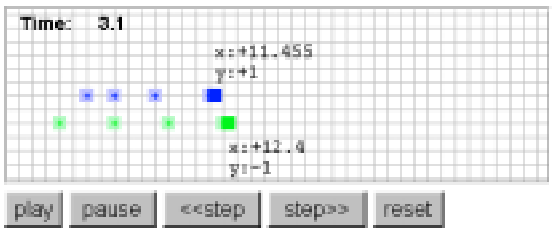
Consider the following distinct forces:

1. A downward force of gravity.
2. A force exerted by the rope pointing from A to O.
3. A force in the direction of the boy's motion.
4. A force pointing from O to A.

Which of the above forces is (are) acting on the boy when he is at position A?

- (A) 1 only.
- (B) 1 and 2.
- (C) 1 and 3.
- (D) 1, 2 and 3.
- (E) 1, 3, and 4.

Question 19

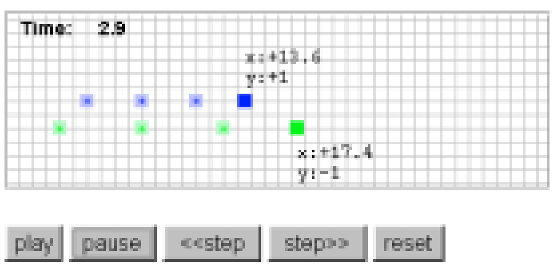


Two blocks are moving as shown in the animation. Do the blocks ever have the same speed?

Start Animation

- A) No
- B) Yes, at t=1s
- C) Yes, at t=4s
- D) Yes, at t=1s and t=4s
- E) Yes, at some time during interval 2s to 3s

Question 20



Start Animation

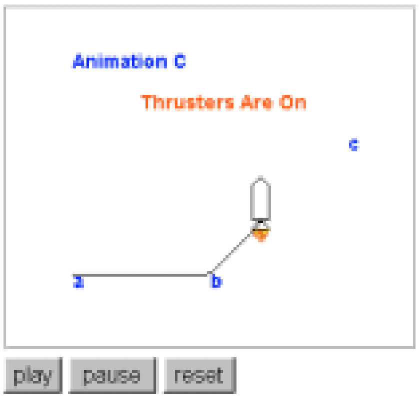
Two blocks are moving as shown in the animation. The acceleration of the blocks are related as follows:

- (A) The acceleration of "blue" is greater than the acceleration of "green".
- (B) The acceleration of "blue" equals the acceleration "green". Both accelerations are greater than zero.
- (C) The acceleration of "green". is greater than the acceleration of "blue".
- (D) The acceleration of "blue" equals the acceleration of "green". Both accelerations are zero.
- (E) Not enough information is given to answer the question.

USE THE STATEMENT AND ANIMATIONS BELOW TO ANSWER THE NEXT FOUR QUESTIONS (21 through 24).

A rocket drifts sideways in outer space from point "a" to point "b" as shown in the animations. The rocket is subject to no outside forces. Starting at position "b", the rockets's engine is turned on and produces a constant thrust (force on the rocket) at right angles to the line "ab". The constant thrust is maintained until the rocket reaches a point "c" in space.

Question 21



Which animation best represents the path of the rocket between points "b" and "c"?

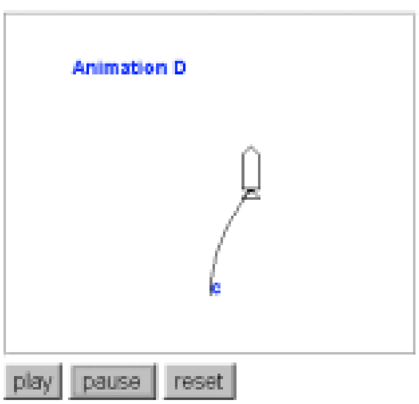
- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 22

As the rocket moves from position "b" to position "c" its speed is:

- (A) constant.
- (B) continuously increasing.
- (C) continuously decreasing.
- (D) increasing for a while and constant thereafter.
- (E) constant for a while and decreasing thereafter.

Question 23



At point "c" the rocket's engine is turned off and the thrust immediately drops to zero. Which animation best represents the path the rocket will follow beyond point "c"?

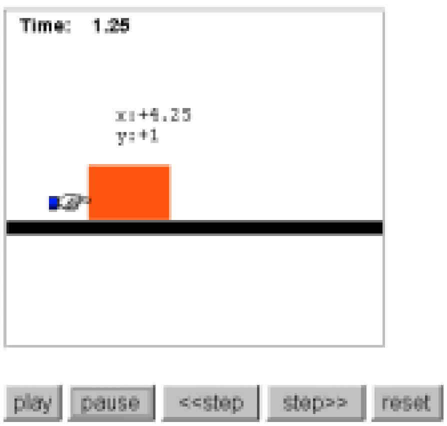
- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 24

Beyond position "c" the speed of the rocket is:

- (A) constant.
- (B) continuously increasing.
- (C) continuously decreasing.
- (D) increasing for a while and constant thereafter.
- (E) constant for a while and decreasing thereafter.

Question 25



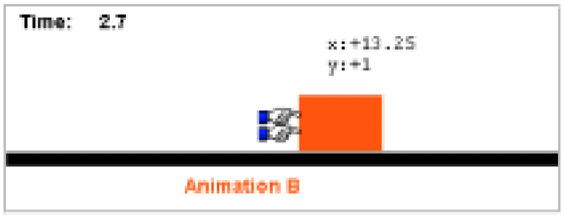
A woman exerts a constant horizontal force on a large block as shown in the animation. As a result, the box moves across a horizontal floor with a constant speed v_o

Start Animation

The constant horizontal force applied by the woman:

- (A) has the same magnitude as the weight of the box.
- (B) is greater than the weight of the box
- (C) has the same magnitude as the total force that resists the motion of the box.
- (D) is greater than the total force that resists the motion of the box.
- (E) is greater than either the weight of the box or the total force that resists its motion.

Question 26

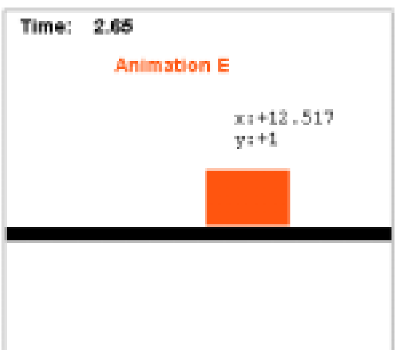


play pause <<step step>> reset

If the woman in the previous question doubles the constant horizontal force that she exerts on the box at t=2 seconds, which animation shows the correct way that the box then moves?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

Question 27

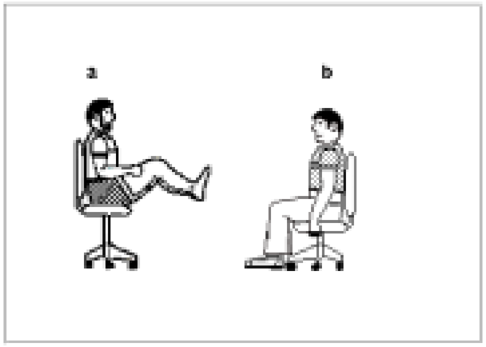


If the woman in question 25 suddenly stops applying the horizontal force at t=2 seconds, which animation shows the correct way that the box moves?

- Start Animation A
- Start Animation B
- Start Animation C
- Start Animation D
- Start Animation E

play pause <<step step>> reset

Question 28



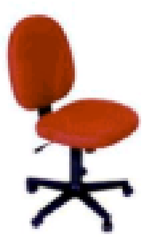
play pause <<step step>> reset

Start Animation

In the animation, student "a" has a mass of 95 kg and student "b" has a mass of 77 kg. They sit in identical office chairs facing each other. Student "a" places his bare feet on the knees of student "b", as shown. Student "a" then suddenly pushes outward with his feet, causing both chairs to move. During the push and while the students are still touching one another:

- (A) neither student exerts a force on the other.
- (B) student "a" exerts a force on student "b", but "b" does not exert any force on "a".
- (C) each student exerts a force on the other, but "b" exerts the larger force.
- (D) each student exerts a force on the other, but "a" exerts the larger force.
- (E) each student exerts the same amount of force on the other.

Question 29



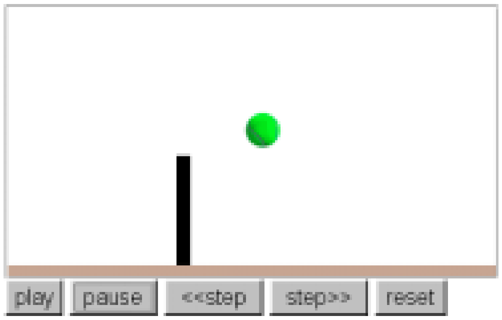
An empty office chair is at rest on a floor. Consider the following forces:

1. A downward force of gravity.
2. An upward force exerted by the floor.
3. A net downward force exerted by the air.

Which of the forces is (are) acting on the office chair?

- (A) 1 only.
- (B) 1 and 2.
- (C) 2 and 3.
- (D) 1, 2, and 3.
- (E) none of the forces. (Since the chair is at rest there are no forces acting upon it.)

Question 30



Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court. Start Animation

Consider the following forces:

1. A downward force of gravity.
2. A force by the "hit".
3. A force exerted by the air.

Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?

- (A) 1 only.
- (B) 1 and 2.
- (C) 1 and 3.
- (D) 2 and 3.
- (E) 1, 2, and 3

When you are finished press the button below to submit your answers. It may take a few minutes for your answers to be processed. Please do not press the button twice. A page will appear notifying you that your answers have been received.

Submit Answers

Appendix B – The Traditional Force Concept Inventory

1. Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. The time it takes the balls to reach the ground below will be:

(A) about half as long for the heavier ball as for the lighter one.

(B) about half as long for the lighter ball as for the heavier one.

(C) about the same for both balls.

(D) considerably less for the heavier ball, but not necessarily half as long.

(E) considerably less for the lighter ball, but not necessarily half as long.

2. The two metal balls of the previous problem roll off a horizontal table with the same speed. In this situation:

(A) both balls hit the floor at approximately the same horizontal distance from the base of the table.

(B) the heavier ball hits the floor at about half the horizontal distance from the base of the table than does the lighter ball.

(C) the lighter ball hits the floor at about half the horizontal distance from the base of the table than does the heavier ball.

(D) the heavier ball hits the floor considerably closer to the base of the table than the lighter ball, but not necessarily at half the horizontal distance.

(E) the lighter ball hits the floor considerably closer to the base of the table than the heavier ball, but not necessarily at half the horizontal distance.

3. A stone dropped from the roof of a single story building to the surface of the earth:

(A) reaches a maximum speed quite soon after release and then falls at a constant speed thereafter.

(B) speeds up as it falls because the gravitational attraction gets considerably stronger as the stone gets closer to the earth.

(C) speeds up because of an almost constant force of gravity acting upon it.

(D) falls because of the natural tendency of all objects to rest on the surface of the earth.

(E) falls because of the combined effects of the force of gravity pushing it downward and the force of the air pushing it downward.

4. A large truck collides head-on with a small compact car. During the collision:

(A) the truck exerts a greater amount of force on the car than the car exerts on the truck.

(B) the car exerts a greater amount of force on the truck than the truck exerts on the car.

(C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.

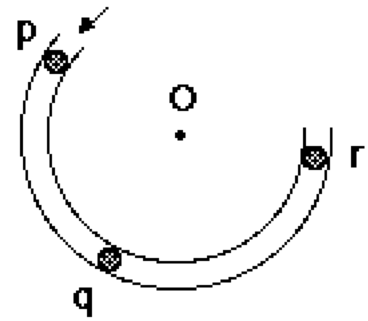
(D) the truck exerts a force on the car but the car does not exert a force on the truck.

(E) the truck exerts the same amount of force on the car as the car exerts on the truck.

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT TWO QUESTIONS

(5 and 6).

The accompanying figure shows a frictionless channel in the shape of a segment of a circle with center at "O". The channel has been anchored to a frictionless horizontal table top. You are looking down at the table. Forces exerted by the air are negligible. A ball is shot at high speed into the channel at "p" and exits at "r."

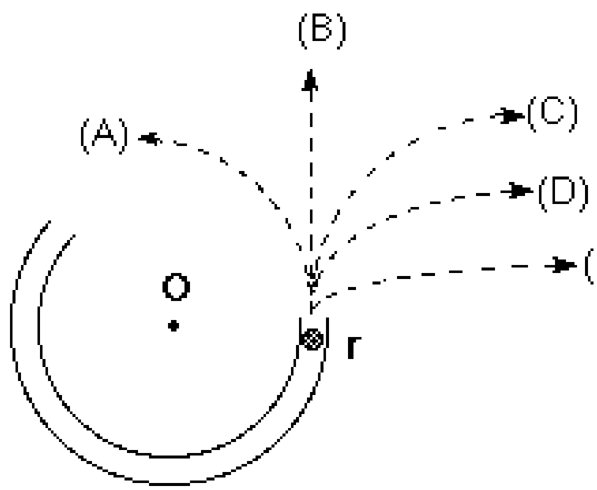


5. Consider the following distinct forces:
    1. A downward force of gravity.
    2. A force exerted by the channel pointing from q to O.
    3. A force in the direction of motion.
    4. A force pointing from O to q.

    Which of the above forces is (are) acting on the ball when it is within the frictionless channel at position "q"?

    (A) 1 only.
    (B) 1 and 2.
    (C) 1 and 3.
    (D) 1, 2, and 3.
    (E) 1, 3, and 4.

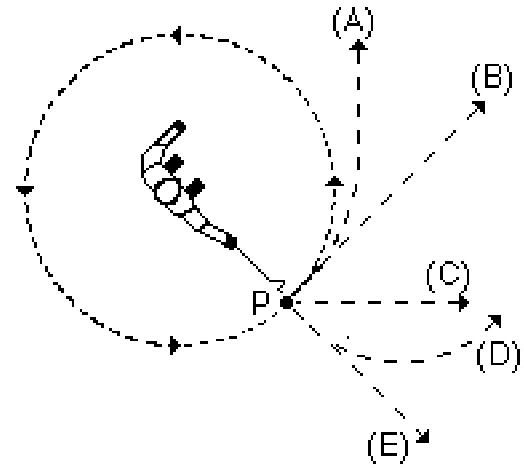
6. Which path in the figure at right would the ball most closely follow after it exits the channel at "r" and moves across the frictionless table top?



7. A steel ball is attached to a string and is swung in a circular path in a horizontal plane as illustrated in the accompanying figure.

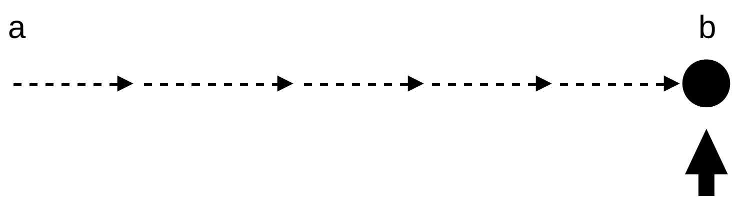
At the point P indicated in the figure, the string suddenly breaks near the ball.

If these events are observed from directly above as in the figure, which path would the ball most closely follow after the string breaks?

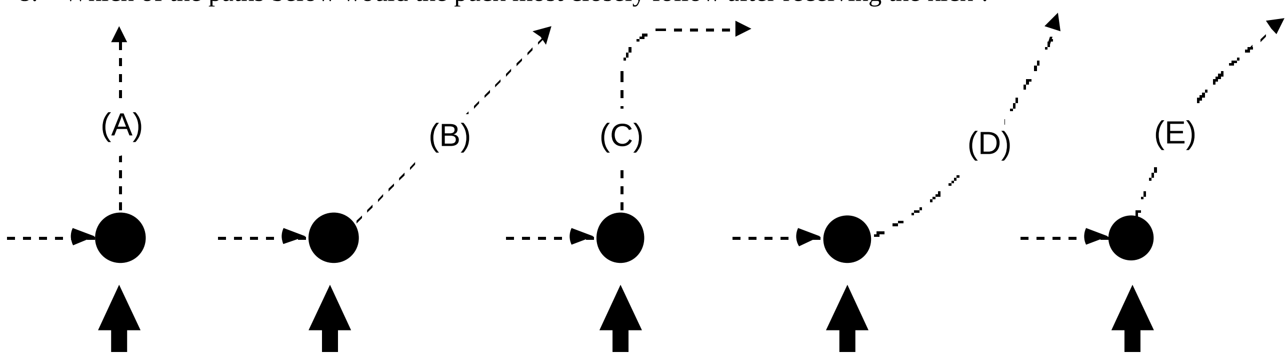


USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (8 through 11).

The figure depicts a hockey puck sliding with constant speed v_o in a straight line from point "a" to point "b" on a frictionless horizontal surface. Forces exerted by the air are negligible. You are looking down on the puck. When the puck reaches point "b," it receives a swift horizontal kick in the direction of the heavy print arrow. Had the puck been at rest at point "b," then the kick would have set the puck in horizontal motion with a speed v_k in the direction of the kick.



8. Which of the paths below would the puck most closely follow after receiving the kick ?



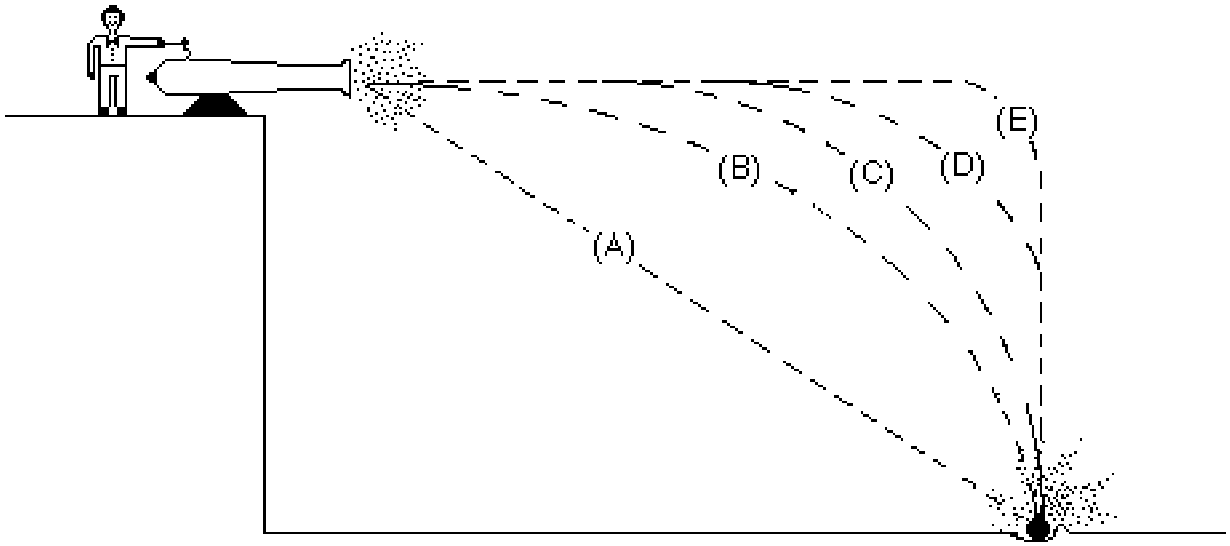
9. The speed of the puck just after it receives the kick is:
   - (A) equal to the speed "v_o" it had before it received the kick.
   - (B) equal to the speed "v_k" resulting from the kick and independent of the speed "v_o".
   - (C) equal to the arithmetic sum of the speeds "v_o" and "v_k".
   - (D) smaller than either of the speeds "v_o" or "v_k".
   - (E) greater than either of the speeds "v_o" or "v_k", but less than the arithmetic sum of these two speeds.

10. Along the frictionless path you have chosen in question 8, the speed of the puck after receiving the kick:
    - (A) is constant.
    - (B) continuously increases.
    - (C) continuously decreases.
    - (D) increases for a while and decreases thereafter.
    - (E) is constant for a while and decreases thereafter.

11. Along the frictionless path you have chosen in question 8, the main force(s) acting on the puck after receiving the kick is (are):
    - (A) a downward force of gravity.
    - (B) a downward force of gravity, and a horizontal force in the direction of motion.
    - (C) a downward force of gravity, an upward force exerted by the surface, and a horizontal force in the direction of motion.
    - (D) a downward force of gravity and an upward force exerted by the surface.

(E) none. (No forces act on the puck.)

12. A ball is fired by a cannon from the top of a cliff as shown in the figure below. Which of the paths would the cannon ball most closely follow?



13. A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):

(A) a downward force of gravity along with a steadily decreasing upward force.

(B) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth.

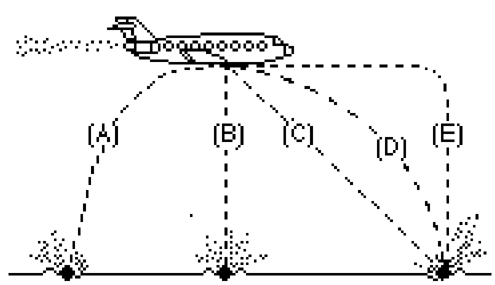
(C) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity.

(D) an almost constant downward force of gravity only.

(E) none of the above. The ball falls back to ground because of its natural tendency to rest on the surface of the earth.

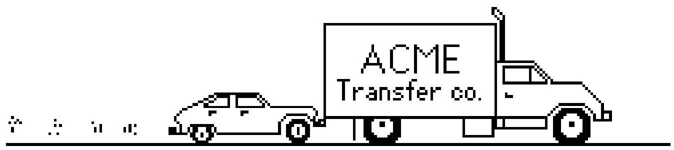
14. A bowling ball accidentally falls out of the cargo bay of an airliner as it flies along in a horizontal direction.

As observed by a person standing on the ground and viewing the plane as in the figure at right, which path would the bowling ball most closely follow after leaving the airplane?



USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT TWO QUESTIONS (15 and 16).

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure below.



15. While the car, still pushing the truck, is speeding up to get up to cruising speed:
    - (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
    - (B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.
    - (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
    - (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
    - (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

16. After the car reaches the constant cruising speed at which its driver wishes to push the truck:
    - (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
    - (B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.
    - (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
    - (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
    - (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

17. An elevator is being lifted up an elevator shaft at a constant speed by a steel cable as shown in the figure below. All frictional effects are negligible. In this situation, forces on the elevator are such that:

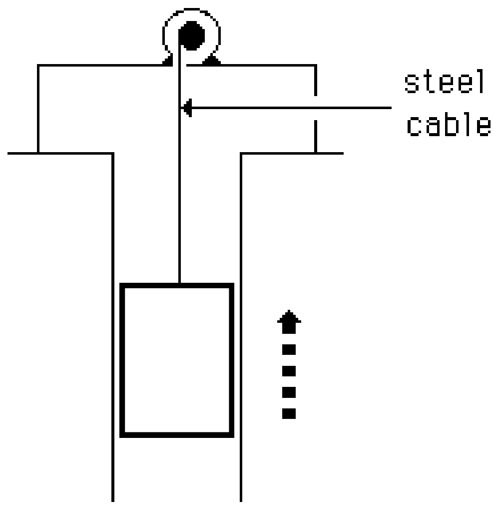
(A) the upward force by the cable is greater than the downward force of gravity.

(B) the upward force by the cable is equal to the downward force of gravity.

(C) the upward force by the cable is smaller than the downward force of gravity.

(D) the upward force by the cable is greater than the sum of the downward force of gravity and a downward force due to the air.

(E) none of the above. (The elevator goes up because the cable is being shortened, not because an upward force is exerted on the elevator by the cable).

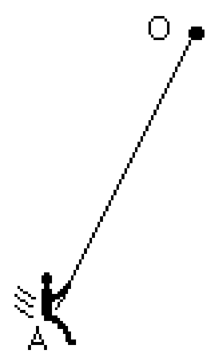


Elevator going up
at constant speed

18. The figure below shows a boy swinging on a rope, starting at a point higher than A.

Consider the following distinct forces:

1. A downward force of gravity.
2. A force exerted by the rope pointing from A to O.
3. A force in the direction of the boy's motion.
4. A force pointing from O to A.



Which of the above forces is (are) acting on the boy when he is at position A?

(A) 1 only.

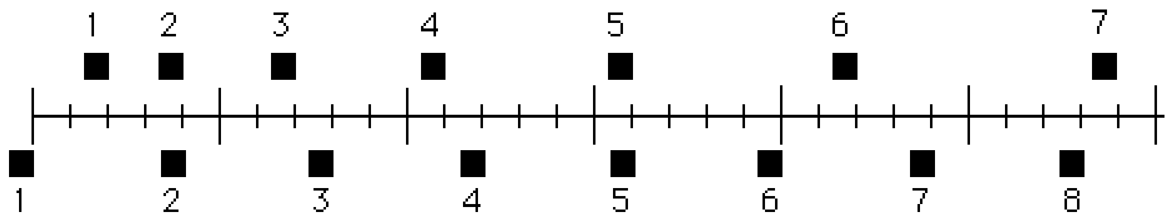
(B) 1 and 2.

(C) 1 and 3.

(D) 1, 2, and 3.

(E) 1, 3, and 4.

19. The positions of two blocks at successive 0.20-second time intervals are represented by the numbered squares in the figure below. The blocks are moving toward the right.



Do the blocks ever have the same speed?

(A) No.

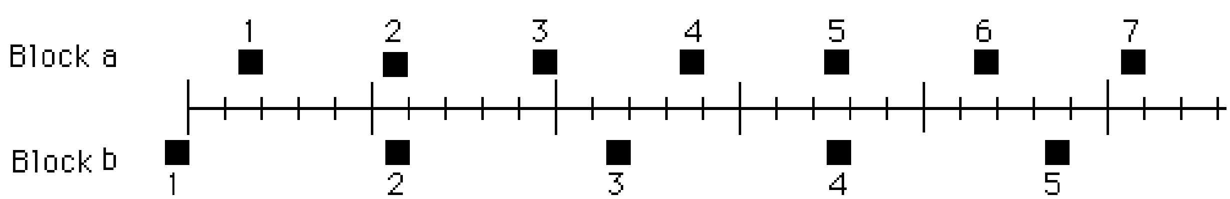
(B) Yes, at instant 2.

(C) Yes, at instant 5.

(D) Yes, at instants 2 and 5.

(E) Yes, at some time during the interval 3 to 4.

20. The positions of two blocks at successive 0.20-second time intervals are represented by the numbered squares in the figure below. The blocks are moving toward the right.



The accelerations of the blocks are related as follows:

(A) The acceleration of "a" is greater than the acceleration of " b".

(B) The acceleration of "a" equals the acceleration of "b". Both accelerations are greater than zero.

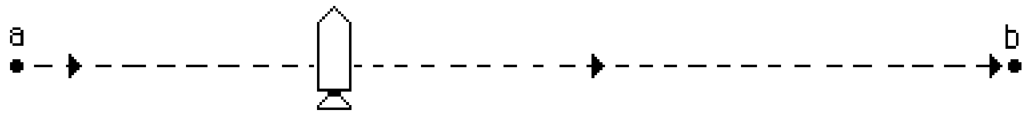
(C) The acceleration of "b" is greater than the acceleration of "a".

(D) The acceleration of "a" equals the acceleration of "b". Both accelerations are zero.

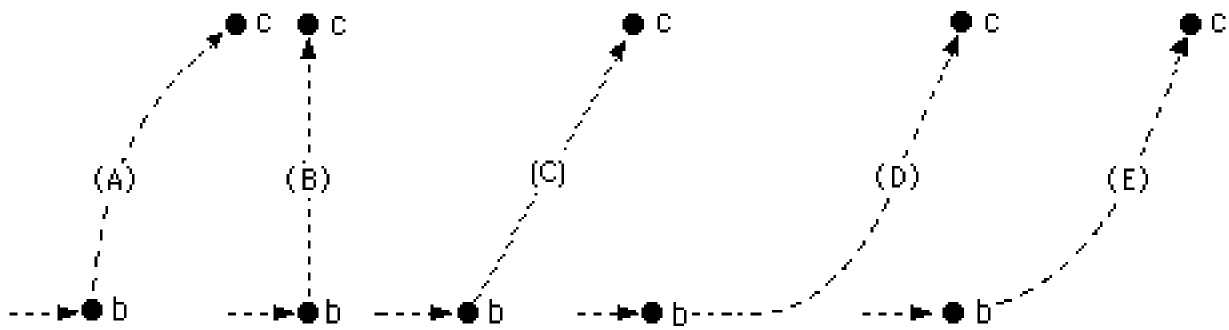
(E) Not enough information is given to answer the question.

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (21 through 24).

A rocket drifts sideways in outer space from point "a" to point "b" as shown below. The rocket is subject to no outside forces. Starting at position "b", the rocket's engine is turned on and produces a constant thrust (force on the rocket) at right angles to the line "ab". The constant thrust is maintained until the rocket reaches a point "c" in space.



21. Which of the paths below best represents the path of the rocket between points "b" and "c"?



22. As the rocket moves from position"b" to position "c" its speed is:

(A) constant.

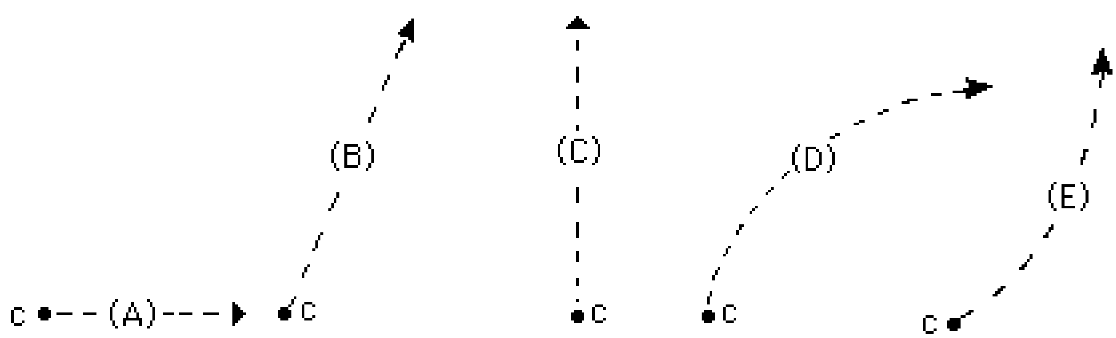
(B) continuously increasing.

(C) continuously decreasing.

(D) increasing for a while and constant thereafter.

(E) constant for a while and decreasing thereafter.

23. At point "c" the rocket's engine is turned off and the thrust immediately drops to zero. Which of the paths below will the rocket follow beyond point "c"?



24. Beyond position "c" the speed of the rocket is:

(A) constant.

(B) continuously increasing.

(C) continuously decreasing.

(D) increasing for a while and constant thereafter.

(E) constant for a while and decreasing thereafter.

25. A woman exerts a constant horizontal force on a large box. As a result, the box moves across a horizontal floor at a constant speed "v_o".

The constant horizontal force applied by the woman:

(A) has the same magnitude as the weight of the box.

(B) is greater than the weight of the box.

(C) has the same magnitude as the total force which resists the motion of the box.

(D) is greater than the total force which resists the motion of the box.

(E) is greater than either the weight of the box or the total force which resists its motion.

26. If the woman in the previous question doubles the constant horizontal force that she exerts on the box to push it on the same horizontal floor, the box then moves:

(A) with a constant speed that is double the speed "v_o" in the previous question.

(B) with a constant speed that is greater than the speed "v_o" in the previous question, but not necessarily twice as great.

(C) for a while with a speed that is constant and greater than the speed "v_o" in the previous question, then with a speed that increases thereafter.

(D) for a while with an increasing speed, then with a constant speed thereafter.

(E) with a continuously increasing speed.

27. If the woman in question 25 suddenly stops applying a horizontal force to the block, then the block will:

(A) immediately come to a stop.

(B) continue moving at a constant speed for a while and then slow to a stop.

(C) immediately start slowing to a stop.

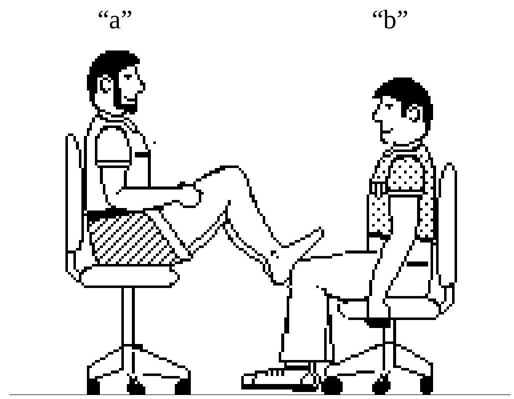
(D) continue at a constant speed.

(E) increase its speed for a while and then start slowing to a stop.

28. In the figure at right, student "a" has a mass of 95 kg and student "b" has a mass of 77 kg. They sit in identical office chairs facing each other.

Student "a" places his bare feet on the knees of student "b", as shown. Student "a" then suddenly pushes outward with his feet, causing both chairs to move.

During the push and while the students are still touching one another:



(A) neither student exerts a force on the other.

(B) student "a" exerts a force on student "b", but "b" does not exert any force on "a".

(C) each student exerts a force on the other, but "b" exerts the larger force.

(D) each student exerts a force on the other, but "a" exerts the larger force.

(E) each student exerts the same amount of force on the other.

29. An empty office chair is at rest on a floor. Consider the following forces:

1. A downward force of gravity.
2. An upward force exerted by the floor.
3. A net downward force exerted by the air.

Which of the forces is (are) acting on the office chair?

(A) 1 only.

(B) 1 and 2.

(C) 2 and 3.

(D) 1, 2, and 3.

(E) none of the forces. (Since the chair is at rest there are no forces acting upon it.)

30. Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court.

Consider the following forces:

1. A downward force of gravity.
2. A force by the "hit".
3. A force exerted by the air.

Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?

(A) 1 only.

(B) 1 and 2.

(C) 1 and 3.

(D) 2 and 3.

(E) 1, 2, and 3.

Appendix C - Consent Form

North Carolina State University
INFORMED CONSENT FORM

Principle Investigator: Melissa H. Dancy Faculty Sponsor: Robert J. Beichner

You are invited to participate in a research study. The purpose of this study is to investigate students' ideas about physics.

INFORMATION
You will be asked to answer a 30 item multiple-choice test related to PY205 topics. You will either be asked to think-aloud as you answer these questions or to discuss your answers with the interviewer after you have completed the test. A video camera and audio recorder will be used to record what you say and do while solving the problems.

RISKS
We do not foresee any risks associated with your participation in this research study.

BENEFITS
You will be paid $10 upon completion of the interview. It is also likely that you will learn a little physics and contribute to the physics instruction of the future.

CONFIDENTIALITY
The data collected will be kept strictly confidential. Data will be stored securely and will be made available only to persons conducting the study unless you specifically give permission in writing to do otherwise. No reference will be made in oral or written reports which could link you to the study. Video tapes will be erased at the conclusion of the study.

CONTACT
If you have questions at any time about the study or the procedures, you may contact the researcher, Melissa H. Dancy, at NCSU's Physics Department, or 515-7059. If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact the Dr. Gary A. Mirka, Chair of the NCSU IRB for the Use of Human Subjects in Research Committee, Box 7906, NCSU Campus.

PARTICIPATION
Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

CONSENT
I have read and understand the above information. I have received a copy of this form. I agree to participate in this study.

Subject's signature____ Date _________________

Investigator's signature___________________________________ Date _________________

Parent's signature (if student is under 18) _____________________ Date _________________

Appendix D - Background Questionnaire

Background Questionnaire

This information will only be used in aggregate form to determine general characteristics of those who were interviewed. It will be held confidential.

Name ______________________________________ Date ___________

Age __________ Gender___________ Class __________ Major ___________

Have you ever taken a physics class before? __________

SAT math score ________ SAT verbal score________ Grade in MA 141_________

Appendix E - Sample Transcript

Tanner 03/30/00
Form A

Question 1 - Animation
T: (Reading Question) Well I think that if air resistance is negligible, everything falls at the same rate, so they should hit the ground at the same time. (Play Animation A) So I'm thinking this is not how I want it because it doesn't go along with what I said.
M: Ok
T: (Plays Animation B) Probably not that one either (Plays Animation C) Alright, that one's looking like what I'd pick but I'm going to check the others. (Plays Animation D) (Plays Animation E) alright I'm gonna go with animation C, do I just click on the circle here?
M: You don't even have to do that? How do you know that, you said that you know they fall at the same rate? How do you know that?
T: Well, I learned that in class, I guess.
M: Ok
T: I think why I knew that, I guess that's just kinda common knowledge, you pick up but, well not really common knowledge but you hear people like say that, like what do you think is gonna fall faster? A pebble or a bowling ball? Then they prove to you that they fall at the same speed.
M: Ok
T: If that's right? (laughs)
Question 2 - Traditional
T: (Reading Question) Well, they're going at the same speed so they have, I guess the same horizontal speed and they have the same vertical speed. So I'd say they both hit the floor at approximately the same horizontal distance.
M: Which choice was that?
T: That was A
M: Ok
Question 3 - Traditional
T: (Reading Question) I think it's uh, has to do with the gravitational attraction. The closer you are to the earth, the greater your gravitational attraction, and it speeds up because, acceleration of gravity, is like 9.8 meters per second. So I'll have to go with B.
M: Ok
Question 4 - Animation
T: (Reading Question) (Starts to activate animation, but does not, goes back to reading question) (Plays Animation) Alright, I'm thinking about, one of Newton's laws that says for every action, there's an equal and opposite reaction. So I'm thinking that, the truck exerts the same amount of force on the car as the car exerts on the truck. So E.
M: Ok
Questions 5 - Animation
T: Alright, should I read all of this? (Reading Question) (Plays Animation A) (long pause) (Plays Animation B) Are these animations for six?
M: Yeah, they are, it gets a little confusing.

T: Ok, alright if the ball is on the earth's surface, I guess there's always a downward force of gravity acting on it. (long pause) I'm not really sure what else to say. (pause) I guess that the force is pointing outwards from the circle. (pause) So lets say 1 and 4. Or that's not even a right answer. (Laughs) (long pause).

M: What are you thinking?

T: I'm trying to think of, like a question I had on my test that had to deal with this, but that was like a month ago, (pause) and I'm pretty sure 1 is one of the options. A downward force of gravity. (pause) I bet there's probably also a radial force on it (Long pause). I'd say E.

M: Ok, so tell me why force one and why forces three and four.

T: Well I was thinking about like when you have like, I guess I played the animations too soon, but I know like when you swing something around it like, it gets pushed out. So I was thinking there's a force pointing zero or from the center out to the edge and I guess that figure there's a force in the direction of the motion.

M: Ok

T: I'm not real sure about that.

M: So E is the answer.

T: Yeah E is…

M: Ok

Question 6 - Animation

T: (Reading Question) (Clicks answer choice, instead of activation link)

M: You have to click up here.

T: Oh yeah. (Plays Animation A) (Plays Animation B) (Plays Animation C)

M: So what do you think about each one as you see it?

T: Well, I think that when you have this, I guess radial acceleration or radial velocity, like when it stops going around in a circle, it just goes straight. So I'm looking for one that just as soon as like this channel ends it starts going straight.

M: Ok

T: I think that's B but I'm going to check the rest of them. (Plays Animation D) (Plays Animation E) (Plays Animation C) Yeah, I'm pretty sure it doesn't (Plays Animation B) go off to the right. I'm gonna go with B, on that one.

M: Ok, and how do you know that it goes straight?

T: (pause) Well, A (Plays Animation A) kind of goes back around, keeps going in a kind of radial motion, and I'm pretty sure it doesn't do that. Cause you think about like an old fashion sling shot or something, like as soon as you… like a David and Goliath sling shot, as soon as you release the stone it keeps going straight.

M: Ok

T: That's kinda my thinking behind that.

Question 7 - Traditional

T: (Reading Question) Alright, then again this is pretty much the same thing. I think its gonna go in the direction of B.

M: Ok

Question 8 - Animation

T: (Reading Question) (Plays Animation A) (Plays Animation B)

M: So what are you thinking as you view the animation?

T: I'm thinking about the puck is like a soccer ball, cause I'm a soccer player.
M: Ok
T: So, it's kind of like (Plays Animation C) visualizing myself kicking the ball. But that's probably not it, because I don't see why would it just make (Plays Animation D) a turn like that. Unless you put some weird spin on it. (Plays Animation E) I think that I'm gonna go (Plays Animation A) with A cause I can just, I can visualize a ball coming towards me from there and then me just kicking it straight.
M: OK
Question 9 - Animation
T: (Reading Question) (looks back at statement that goes with this set of problems) Alright, V_k is the velocity of the kick right? So taking that V_k is the direction of the kick, then I'd probably say equal to the speed V_k resulting from the kick and independent of the speed Vo cause its in different direction.
M: Ok
T: I guess the horizontal velocity has nothing to do with the vertical velocity.
Question 10 - Animation
T: (Reading Question) Well, I'm thinking that the soccer ball would slow down, but that's probably because of friction, and something in motion stays in motion, unless a force is acting on it. So I think that the speed would be constant.
M: Ok
T: Providing it's a frictionless path.
Question 11 - Animation
T: (Reading Questions) Well, I guess there's a downward force for gravity cause then again it's on the earth and everything on the earth is affected by gravity and I guess on the upward force exerted by the surface would be a normal force. I'd go with D.
M: Ok, and why not any of the other forces?
T: Well, after it receives the kick there's.. it goes in a vertical (?) it's in a vertical direction, right? (long pause – reading general statement again) ok, its going horizontal right? and it goes horizontal again. (long pause) This one is kind of tricky.
M: What are you trying to figure out?
T: I was thinking that, the puck slides across the horizontal surface, and then gets kicked in a horizontal motion. So its, I was thinking that it would go in a horizontal motion across here and then go in a vertical motion here, but they're saying it's a horizontal motion after it's kicked. So if it's in a horizontal motion after it receives the kick then there's no other force acting on it cause, it stays… an object in motion stays in motion unless acted on by a force. So its already received the horizontal force, so now its just gravity pushing it down and the surface pushing it up.
M: Ok
Question 12 - Animation
T: (Reading Question) (Plays Animation A) (Plays Animation B) Right now I'm thinking its probably gonna be, … that seems a lot more realistic to me than this one right here. (Plays Animation A)
M: Ok
T: Cause its going horizontally, its not just, it looks like it'd be shot out a the cannon facing this angle right here (points to angle between horizontal and vertical) shooting

down, (Plays Animation C) so I'm, and I know that's probably not the answer, cause as soon as it leaves the cannon its being affected by gravity.
M: Ok
T: (Plays Animation D) That's probably not it, C's that's not it. (Plays Animation E) Yeah that's definitely not it, so I'll go with animation B.
M: Ok
Question 13 - Animation
T: (Reading Question) I'm thinking its D cause it's the force of, being thrown upward is, its already had that force acting on it so its just going up. And then the force, the gravity is continuously pulling it down, so it finally stops it from it from going up and brings it down. So I don't think its none of the above. So I'm gonna go with D. (Never plays Animation)
M: Ok
Question 14 - Traditional
T: (Reading Question) Alright, right here, I'll say D cause the bowling ball is going the same speed as the plane is because its in the plane.
M: Ok
T: And it follows the same path as that like the cannon ball did in the earlier problem.
M: Ok
T: Because of the force of gravity, so I'll say D.
Question 15 - Traditional
T: (Reading Question) Right now, I'm thinking that might be a good possibility cause of the equal and opposite force, The amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car - no I don't think it's that one. The amount of force which the car pushes on the truck is greater than that with which the truck pushes on the car - I don't think it's that one. The car's engine is running so the car pushes against the truck … I'm pretty sure that's not it cause the equal and opposite reaction. Neither car no truck exert any force on the other - I'm pretty sure there's forces acting there someplace (?) Ok I'd say the amount, I'd say A.
M: A?
T: Yeah
Question 16 - Traditional
T: (Reading Question) (Reads choice A) I'm still pretty sure its that one, cause I don't think cruising speed has to do with anything, has to, makes a difference. Cause of Newton's law.
M: Ok
T: The amount of force… yeah I think its A for that one too.
Question 17 - Traditional
T: (Reading Question) (Reads Choice A)At first I'm thinking that that's a good possibility cause the upward force is defying the gravitational law, and pulling it up at the same time. So I'm thinking the upward force is greater than the gravity. (Reads Choice B) I don't think it's that one, cause if it was, then it would just be standing still. (Reads Choice C) No, cause then it wouldn't get off the ground. (Reads Choice D) Well it says that all frictional effects are negligible, so I think air resistance would be a negligible force. (Reads Choice E) I'd go with A.

M: Ok

Question 18 - Traditional

T: (Reading Question) Well, (pause) the question before this was kind of similar to this. I think it was the ball on the track, whatever. But I'm having second thoughts on that, cause I know there has to be a force pulling the boy down, which is gravity cause there's a downward force of gravity. But I also know that A is also pushing the boy up. So that would probably be from A to zero… and there's probably a force in the direction of the boy's motion, cause he's going back and forth.

M: Ok

T: So I'd say, so that's one of the answers. 1, 2, & 3. D.

M: Ok

Question 19 - Animation

T: (Reading Question) (Plays Animation) I'd have to see that one again. (Plays Animation) (Plays Animation) (Long Pause)

M: What are you thinking?

T: Well I know that, speed is velocity, which is change of distance over change of time. I'm trying to figure out if when they're at the same point had they traveled the same distance over the same amount of time. I guess that that instant they'd have the same instantaneous velocity. (Plays Animation) (?)…(Plays Animation) (Plays Animation) It looks like they're moving the same velocity at one time. It looks like they're not the same point, but they're covering the same ground in the same amount of time, the same distance. (Plays Animation) In right about in there. (Points to moving blocks) So I guess that'd be from 2 to 3 seconds. So E.

M: Ok

Question 20 - Traditional

T: (Reading Question) alright, I think that, even though B is going faster than A, they're both at constant velocities which would mean they'd have, zero accelerations both of them so their accelerations would both equal zero. So I'd say D

M: Ok

T: Cause they look like they don't, … they look equal distances

M: Ok

Question 21 - Traditional

T: (Reading Question) (Long Pause)

M: What are you thinking?

T: Constant thrust will probably give it a constant velocity, and this (points to "C") is looking kind of like a constant velocity graph.. the slope of a graph, the slope of a position time graph is velocity and that looks like it has a constant slope to it. So I'd say C.

M: Ok

Question 22 - Traditional

T: (Reading Question) This is basically just saying what I just went over, where the, looks like a curve to a position time graph, where slope is the velocity. So, I think it'd be constant. So A.

Question 23 - Traditional
T: (Reading Question) Alright, since its going up in the air, and its still going at like, I guess a kind of horizontal direction, its not gonna, after C its not gonna keep going at the same vertical velocity. It won't, probably won't go straight up and this (points to "E") is kind of where it's on an exponetial curve where it keep, where it's more like an acceleration curve. So I think its that. So I think its D, cause it looks like force will start bringing it back down.
M: Ok
T: Or gravity.
Question 24 - Traditional
T: (Reading Question) I'm going to read the questions, see if there's anything about air resistance. I think it's gonna be that the thrust is cut off so it spent longer going at a constantly velocity. So it'll continuously decrease, because there's no more power to keep it at that constant velocity.
M: Ok.
Question 25 - Animation
T: (Reading Question) (Plays Animation) Right now I think its,(?) … is greater than either the weight of the box for a total force that resists… Ok, I think this all has to do with, I guess horizontal motion and I don't think that the woman pushing the box has anything to do with the weight of the box. Cause she might be able to push a hundred-pound box, but she wouldn't be able to pick it up. So I think its greater than the total force that resists the motion of the box, cause if they were equal it'd probably just stay still. But if it were greater then it'd start moving forth.
M: Ok
T: So I'd say D
M: Ok
Question 26 - Animation
T: (Reading Question) (Plays Animation A) (Plays Animation B) (Plays Animation C) I'm pretty sure if she doubles the force it's gonna move faster. Probably not that cause, probably will keep the same constant velocity after time, 2 seconds. (Plays Animation D) That looks like it keeps constant velocity even though a second force is applied on it so I don't think its that. (Plays Animation E) That looks like it accelerates. I don't think it accelerates except at that instance. (Play Animation A) So I'm thinking A or B. Watch these two again. (Play Animation B) B looks like it remains the same (Plays Animation B) velocity or maybe it speeds up a little bit. No that's kind of hard to tell. (Play Animation A) I'm going to watch A again. (Plays Animation B) I think I'll go with A.
M: Ok
T: Cause it seems like it doubles its speed, with a double and other force.
Question 27 - Animation
T: (Reading Question) (Plays Animation A) I don't think it stops (Plays Animation B) automatically. I think its gonna… it won't go. That might be it, cause its not gonna (Plays Animation C) automatically stop. But, its gonna keep going but slow down. (Plays Animation D) Yeah, B and C are looking… they're reasonable. Probably not that cause its (Plays Animation E) not gonna keep going at the same velocity, cause friction, there's a force resisting. Its not gonna speed up then slow down. (Plays Animation B) (Plays Animation C)Alright I think its C, cause animation B looked like it, after she stopped

pushing, it still started going at the same velocity. And then it just started to slow down all of a sudden. As soon as the lady stopped, see it started to slow, it didn't stop but it started to slow down. So I'm gonna go with C
M: Ok

Question 28 - Traditional

T: (Reading Question) I think its each student exerts the same amount of force on each other, E. Back to the equal and opposite reaction.
M: Ok

Question 29 - Animation

T: (Reading Question) 1 and 2, which would be B, cause (pause) I'm not really sure how much force the air would be. I'm pretty sure it's B.
M: Ok
T: But, right now I'm not sure about that air, cause you think about a plane flying it has the wings produce like a… the air produce the upward force. I don't think the air has much to do with the forces acting on the chair. So I'd say B.
M: Ok

Question 30 - Traditional

T: (Reading Question) Alright I don't think it… back to an object in motion stays in motion. It's already had the horizontal force I guess, as the hit so it no longer has that force. It just has the propulsion from the hit.
M: Ok.
T: So it's not 2. I guess there's a strong wind day. So I guess the air is playing a factor on this strong windy day. So the air is pushing it back and then there's a downward force of gravity always. So C

M: Ok great. What I want you to do now is, I'm gonna pull up a different set of questions here, and these are some that you've already answered, and I just want you to look at them again.
T: Are those the ones that I got wrong or something?
M: No not necessarily, you are still safe here.
T: Ok
M: they're 21 through 27, I'll show you, you'll see what I want you to do here, and what I want you to do is just tell me if you want to change your answers. So you'll see the questions… in a different…. Just wait a minute, the different format. So they're all the same questions and same answers.
T: Yeah
M: I'll let you scroll down there. They just look a little bit different.
T: You want me to go through the same procedure?
M: Yeah or you can just partially read them, you'll probably recognize the questions since you just did them.

<u>Going over the questions</u>

Question 21 - Animation

T: (Reading Question) (Plays Animation A) This is kind of like the soccer ball question I guess. (Plays Animation B) (Plays Animation C) I was thinking of this as like a position time, but I guess that really doesn't have to do anything with that. (Plays Animation C) (Plays Animation D) I don't think that there's a force acting on it as soon as it hits B. (Plays Animation E) I don't know, it might be thrust maintained until it, its constant thrust so I don't think it would... (pause) that one seems a lot trickier now. (pause) Cause I know when… (pause) I think it might be B though, actually cause I think it accelerates upward cause, you think about a rocket taking off then go a hundred miles an hour. Right at the instant the rocket fire, I mean the engine fires. It has to build up to that speed with the constant force, with a constant thrust. So I might go with…, (Plays Animation E) I think its animation E.
M: So why, so that's different than what you answered before. What do you think made you change your answer?
T: The animation, I guess. I was thinking like more of like a position time graph, more than like a, just the actual path that it goes in. I guess that's the reason why I changed my answer.
M: Ok, so the animation helped you to see?
T: Yes, it did.

Question 22 - Animation

T: (Reading Question) And then since I picked animation E, I figured it accelerated. So I think from B to C continuously increases.
M: So change to B?
T: Increases for a while at least. Yeah it continuously increases I believe until the rocket is cut off.
M: Ok, and you're changing that mainly because you changed the first one?
T: Yeah mainly because I changed the first one.
M: Ok.

Question 23 - Animation

T: (Reading Question) (Plays Animation A) (Plays Animation B) (Plays Animation C) I don't think its that one. (Plays Animation D) Maybe D. (Plays Animation E) I don't think it's C cause that looks like it accelerates. (Plays Animation A) A looks like it just stops right there. (Plays Animation B) I don't think its that either. (Plays Animation C) (Plays Animation D) I think it's probably D, because it keeps going up but at a smaller velocity every time. So if I'm not mistaken, this is negative acceleration, cause the velocity, the slope gets smaller and smaller.
M: Ok.
T: So I think that one's D. (Plays Animation D) Yeah D.

Question 24 - Animation

T: (Reading Question) continuously decreasing.
M: Ok

Question 25 - Traditional

T: (Reading Question) I think this one is still greater than the total force, of the resistance of the motion of the box.
M: Ok
T: I'm pretty sure it doesn't have to do much with the weight of the box. Cause, I know I can push a lot of more than what I can lift up, and I don't think,… thinking back to like

some of the problems we did in the class, I don't think we really accounted for the gravitational force cause that's a vertical vector or whatever and this motion is horizontal.

Question 26 - Traditional

T: (Reading Question) Alright I'm sure it's not a continuously increasing speed, cause then it would be accelerating. I don't think it accelerates with a constant speed that is double the speed V_0. I'm not really sure that its double the speed. It's not necessarily doubling the speed because when you're pushing it first, part of your force goes to break the friction or resistance force and then part of your force goes to move the box. So I don't think it's necessarily twice as great so I think it would be B.

M: Ok, so why do you think since now that you've changed your answer, why do you think you changed it?

T: I guess I kind of looked at it different from… you see this one says then with a speed that increases thereafter I guess I wasn't looking at the animation too closely, cause I didn't really recognize that the speed increases. I thought that it just kept at a constant speed once the second force was applied on it.

M: Ok

Question 27 - Traditional

T: (Reading Question) No I don't think it's that one. I think I'll stick with immediately start slowing to a stop.

M: Ok.